

STIC-Biotech/ChemLib

8B13

8E12

From: Davis, Natalie
Sent: Tuesday, August 27, 2002 5:08 PM
To: STIC-Biotech/ChemLib
Subj ct: 09/807470

Please search SEQ ID NO: 1-4 for 09/807470 and interference.

Natalie A.Davis, PhD
Patent Examiner
Art Unit 1642
CM1, Rm 8B13
Mailbox 8E12
Ph (703) 308-6410

Point of Contact
Susan Hanley
Technical Info. Specialist
CM1 6B05 Tel: 305-4053

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

Point of Contact
Susan H. Hsieh
Technical Info Specialist
CMI 8800 Tel. 303-4053

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 04:39:56 ; Search time 5654.61 Seconds
(without alignments)
8552.531 Million cell updates/sec

Title: US-09-807-470-1

Perfect score: 2311

Sequence: 1 agagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score Match	Length	DB ID	Description

1	2266.8	98.1	155623	2	AC106086	AC106086 Rattus no
2	1092	47.3	1092	6	BD005227	BD005227 Method of
3	808.4	35.0	165538	9	AC093805	AC093805 Homo sapi
4	695.2	30.1	1288	6	BD005228	BD005228 Method of
5	647.8	28.0	1110	6	BD005226	BD005226 Method of
6	545.4	23.6	2856	4	CFTRAM	X63678 C.familiari
7	529.2	22.9	2722	9	BC000687	BC000687 Homo sapi
8	493.8	21.4	1267	6	AX329795	AX329795 Sequence
9	493.8	21.4	1267	9	BC005229	BC005229 Method of
10	493.8	21.4	1267	9	HSTRAMP	X63679 H.sapiens m
11	486	21.0	2819	10	BC012401	BC012401 Mus muscu
12	481.6	20.8	2387	4	BTU19578	U19578 Bos taurus
13	479.4	20.7	2720	10	AY029764	AY029764 Mus muscu
14	438.6	19.0	1338	5	AY029534	AY029534 Xenopus l
15	390.8	16.9	1453	5	AY029529	AY029529 Danio rer
16	345	14.9	642	6	AX303127	AX303127 Sequence
17	259.6	11.2	6974	9	HUMORFKG1A	D31762 Human mRNA
18	245.4	10.6	1408	10	BC018212	BC018212 Mus muscu
19	244.8	10.6	1360	10	AY029530	AY029530 Mus muscu
20	220.4	9.5	1272	5	AY029535	AY029535 Xenopus l
21	212.6	9.2	2123	3	AY075393	AY075393 Drosophil
22	212.6	9.2	132910	2	AC014319	AC014319 Drosophil
23	212.6	9.2	146717	3	AC097726	AC097726 Drosophil
24	212.6	9.2	299275	3	AE003419	AE003419 Drosophil
25	211	9.1	212370	3	DMBR7A4	AL109630 Drosophil
26	117.2	5.1	73339	2	AC027246	AC027246 Homo sapi
27	96	4.2	163997	9	AC022731	AC022731 Homo sapi
28	93	4.0	196349	2	AC022032	AC022032 Homo sapi
29	82	3.5	377	11	AU026177	AU026177 Rattus no
30	74.4	3.2	46241	2	AC107646	AC107646 Mus muscu
31	74.2	3.2	196349	2	AC022032	AC022032 Homo sapi
32	73.2	3.2	1694	9	HSM800167	AL049382 Homo sapi
33	70.4	3.0	64530	2	AC110042	AC110042 Mus muscu
34	70.2	3.0	1640	9	AB070034	AB070034 Macaca fa
35	69.2	3.0	192079	2	AC068141	AC068141 Mus muscu
36	69.2	3.0	349980	6	AX344566	AX344566 Sequence
37	68.8	3.0	1419	8	CCCHSMR	Z67988 C.chinensis
38	68.8	3.0	99003	2	AL390756	AL390756 Homo sapi
39	68.4	3.0	793	9	BC018189	BC018189 Homo sapi
40	68.2	3.0	3265	6	AX092338	AX092338 Sequence
41	68	2.9	2131	10	BC017614	BC017614 Mus muscu
42	67.8	2.9	18011	6	AX344936	AX344936 Sequence
43	67.4	2.9	189572	2	AC084812	AC084812 Homo sapi
44	67	2.9	545	6	AX185705	AX185705 Sequence
45	67	2.9	925	3	AEVA0440X	L29571 Aequorea vi

ALIGNMENTS

RESULT	1	AC106086	155623 bp	DNA	linear	HTG 12-JAN-2002
LOCUS		Rattus norvegicus clone CH230-126G18,	***	70 unordered pieces.		SEQUENCING IN PROGRESS
DEFINITION		AC106086				
ACCESSION		AC106086				
VERSION		AC106086.1	GI:18138601			
KEYWORDS		HTG; HTGS_PHASE1.				
SOURCE		Norway rat.				
ORGANISM		Rattus norvegicus				
REFERENCE		1 (bases 1 to 155623)				
AUTHORS		Muzny,D.M., Adams,C., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carro,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,				

```

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzkler,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlecyk,K., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 155623)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHHR
Center clone name: CH230-126G18
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 127470 bases at least Q40
Consensus quality: 134419 bases at least Q30
Consensus quality: 139475 bases at least Q20
Estimated insert size: 122947; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft.data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 5824: contig of 5824 bp in length
* 5825
* 5925: contig of 4369 bp in length
* 10293: gap of unknown length
* 10394
* 13928: contig of 3535 bp in length
* 13929
* 14028: gap of unknown length
* 14029
19438: contig of 5410 bp in length

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

QY 1621 tagagatttctgtcttcagacgggtcgctgtctttttaaattgtaattgtct 1680
DB 3019 TAGAGATTTTCGTTCTTCAGAACGGGTGCTGCTTTTGAATATGCTAATGATGTCT 3078
QY 1681 aatgtgtttttaaggttttcagacgtatgagtgggggtggtggttaagactaaaccac 1740
DB 3079 AATGTGTTTTAAGGTTTTGCAGACGTATGAGTGGGGATGGGGTAAAGACTAAACCAC 3138
QY 1741 tcagcctctaaatacagtcagaaagtttaacgaccacaacatttatttagttagttctt 1800
DB 3139 TCAGCCTCTAAATACGTCTCAGATAGTTTAAACGACCAACATCTTATTTAGTTAGTTCTT 3198
QY 1801 acctcaacgatttccaaagttttgtgtgatgactgcagaaattgtgtacataaataat 1860
DB 3199 ACCTCAACGATTTTCCAAACGTTTTGTTGTGATGACGTCGAGAAATGTGTACATAAATAAT 3258
QY 1861 agtttctctgtccaaatgtttcttatacgaatttaacaagttctgtagcaagtggtttgtt 1920
DB 3259 AGTTTCTGCTTCCAATGTTCTTTATCGAATTAACAAGTCTGCTAGCAAAAGTGGTTGTT 3318
QY 1921 ttctcaattctctcgcaggataaagtggaataatcgtataaaggttaactcaaatcag 1980
DB 3319 TTCTCAATGTTCTCTGCAAGTAAAGTGAATAATCTGATAAGGTTAAACTCAAAATCAG 3378
QY 1981 tattatgtaacgttggtgttttttaaagtggttttaaatttacaatgaaagcatttgt 2040
DB 3379 TATTATGTAACGTTGGGATTTTTTAAAGTGTTTTAAATTTTACAAATGGAAGCAATGCT 3438
QY 2041 caaacaccacaaataatgtgttttaattttatgtagtagtaattgttagtcttcacgcccc 2100
DB 3439 CAAACACCACAAATAATGTGTTTAAATTTATGAGTAGTAATGTTAGTCTTACGCCCCC 3498
QY 2101 attaaagcacaataatgaatagatgacatgtgtgtgatattgacatttagcgaaatca 2160
DB 3499 ATTAAAGCATCAAAATATGAATAGATGACATGTGTGTTGATATGACATTTAGCGGAATCA 3558
QY 2161 agatacctttaa caaataatgtgtgtgttactaaagaagtaaacgacttctctgtttatt 2220
DB 3559 AGATACCTTTAATAATATGTTGGGTACTAAAGTAGTAACGACTCTCTCTGTTATT 3618
QY 2221 ttaaacactgttacaggaactcgcaaaattaaattactactgaaaaaa 2270
DB 3619 TTAACACTGTACAGGAACACTCGAAAAATTAATATTTACTGAAAAAA 3668

RESULT 2
BD005227 1092 bp DNA linear PAT 31-JAN-2002
LOCUS
DEFINITION Method of screening of protein.
ACCESSION BD005227
VERSION BD005227.1 GI:18633188
KEYWORDS JP 03075332-T/2.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1092)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-T 2 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD, NAOKI TODO, HAJIME OKUYAMA, OTOAKI
IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO
OS Rattus sp. (rat)
PN JP 03075332-T/2
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PR 20-AUG-1999 JP 99P 234764
PI NAOKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIRONORI ISHIKAWA,
KIYOMITSU NEMOTO
PC C12Q1/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
CC (C12P21/02,C12R1:91)

FEATURES
source
BASE COUNT 260 a 268 c 288 g 276 t
ORIGIN
Query Match 47.3%; Score 1092; DB 6; Length 1092;
Best Local Similarity 100.0%; Pred. No. 1.9e-209;
Matches 1092; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 463 atggggctccgcaagaagaacccaggaacccccccgggtgctgagccagaatctatggtg 522
DB 1 ATGGGGCTCCGCAAGAAGAACGCCAGGAACCCCGGGTCTGAGCACGAATTCATGTTG 60
QY 523 cagaaccacgagatattgtctccctgctggcgtggcgtgttctctgctggtgacttatgttc 582
DB 61 CAGAACCCAGCGGATATGCTCTCTGCGTGGCATGTTCTTCTGCTGGGACTTATGTTTC 120
QY 583 gagggcacggccgagatgctgatcgtgttctcaccctgcagcagtgaggatcgttgcaca 642
DB 121 GAGGCGACGGCGAGATGTCGATCGTGTCTCACCTGCAGCATGGAGTCGTTGCCCA 180
QY 643 ggggaagggtaccctcggggtccaggaccctttaccattatgggggtcaaaagatctggcc 702
DB 181 GCGGAAGGGCTACCTCGGGGTCCAGGACCCCTTACCATATATGGGGTCAAAAGATCTGCC 240
QY 703 acagtgttcttctacatgctgtgtggccatcatctcagccaccattcagagtcagtg 762
DB 241 ACAGTGTGTTCTTACATGCTGTGTGGCCATCATCATTCAGCCACCATTTCAGAGTACGTG 300
QY 763 ctagataagctcagcggagactgcagctcaccgaaggcaaaacaaatgtaaatgag 822
DB 301 CTAGATAAGCTCAGCGGAGACTGCAGCTCACAAAGGCAAAACAAATTAATGATGAG 360
QY 823 gcgggcagctgagtggttctacatagtgctgtgtatctggtggtatgatactatggcc 882
DB 361 GCGGGCAGCTGAGTGTGTTTACATAGTGTCTGGTATCTGGGGTATGATCATCTTGGCC 420
QY 883 tctgagaactgcctgtcagacccccactctattgtggaagtcctcagccccacacatgatg 942
DB 421 TCTGAGAAGTGCCTGTGCAGACCCCACTCTATTGTGGAAGTCTCAGCCCCACAACTGATG 480
QY 943 acatttcagatgaattttctacatctcaacagttggcttactgtgttcaatgattcccg 1002
DB 481 ACATTTTCAGATGAATTTTTTCTACATCTCACAGTTGGCTTACTGTTTATAGTTTCCCG 540
QY 1003 gagctctacttcagaagaagtcaggaacaagatatccgggttcaactcatcatctggc 1062
DB 541 GAGCTCTACTTCAGAAAGTCAGGAACAAGATATCCCGGGTCACTCATCTACATTTGGC 600
QY 1063 ctccaccttctccacattggagggtccctatctcttctgacttgaacacacctggcctgctg 1122
DB 601 CTCACCTCTTCCACATTGGAGGGCCCTATCTCTTGTACTTGAACACCCTGGCCCTGCTG 660
QY 1123 ctctgtatgctgactatgctgtcagatcctctccagcgtgtgcagcctgttcacttt 1182
DB 661 CTCTGTATGCTGCTACATATGCTGTCAGGTCTCTCAGCGTGTGAGCCCTGCTTTACTTT 720
QY 1183 ggggatgagcgttacagaagaaggtgtcttcttggcctatctgttttatatccggaga 1242
DB 721 GGGGATGAGCGGTACAGAAAGGGTGTCTTTGTGGCCCTATCTGTTTATATATCCGGGAGA 780
QY 1243 ctctgtacactgattgtctcagttgttacagtagggttctcacttggcggggaacaaatcgg 1302
DB 781 CTCGTGACATGATTTGCTCAGTGTGTACAGTAGGGCTTCACTTGGCGGGACAATCGG 840
QY 1303 aatgaaatgctctctctctctggttaattgcaatgtgttggcagctaaatcgtgttctgtcc 1362
DB 841 AATGAAATGCTCTCTCTGGTAATGTCATATGTTGGCAGCTTAAATCGCTTCTTGTCTCC 900

```

QY 1363 tcaggtgcagttccaggtgtacataacatgacottgacacgtctggtcttcagaga 1422
Db 901 TCCAGTTGCAGTATCCAGGTTGATACATAACATGGACCTTGACGACCGTCTGGCTTCAGAGA 960

QY 1423 tggtagaagatcgaaatcttcattgtctgtgaggaagagacggtccaggtcagagaaa 1482
Db 961 TGTTAGAAGATCGGAATCTTCATGTCGTGGGGAAGAGACGGTCCAGGTCGAGAGAAA 1020

QY 1483 ggcacagaaatgagtggaagatccaaatagatagattctccaccacaaagaagaagag 1542
Db 1021 GGCACAGAAATGGAGTGGAGAAATCCAAATAGATATTCTCCACCAAGAAAGAG 1080

QY 1543 aaagctcccttag 1554
Db 1081 AAAGCTCCCTTAG 1092

RESULT 3
AC093805/c
LOCUS AC093805 BAC clone RP11-32609 from 4, complete sequence.
DEFINITION AC093805 AC012532
VERSION AC093805.3 GI:18497223
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 165538)
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99083792
REFERENCE 2 (bases 1 to 165538)
AUTHORS Cedroni,M., Abbott,A. and Bielicki,L.
TITLE The sequence of Homo sapiens BAC clone RP11-32609
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 165538)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 165538)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 165538)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 5, 2002 this sequence version replaced gi:16077043.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0326009
Drafting Center: WIBR
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenio,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://dacpac.med.buffalo.edu>)
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-32609; actual end is at base position 165538 of RP11-32609.

Data from AC013556 was used to finish the clone, AC093805. Polymorphisms have been identified between AC013556 and AC093805.

The sequence of AC012532 has been incorporated into AC093805.

FEATURES

source	Location/Qualifiers
1. .165538	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-32609"
1. .130	/clone_lib="RPCI-11"
	/rpt_family="MaLR"
131. .436	/rpt_family="ERV1"
870. .1245	/rpt_family="L1"
1246. .2720	/rpt_family="L1"
2721. .3067	/rpt_family="L1"
3627. .3713	/rpt_family="MaLR"
4217. .5074	/rpt_family="AchoBo"
5055. .5101	/rpt_family="(TTTA)n"
5075. .5384	/rpt_family="Alu"
5385. .5592	/rpt_family="AchoBo"
7152. .7426	/rpt_family="L1"
7427. .7734	/rpt_family="Alu"
7590. .7619	/rpt_family="AT-rich"
7735. .7872	/rpt_family="L1"
7873. .8314	/rpt_family="ERV1"
8315. .9564	/rpt_family="L1"
9438. .9470	/rpt_family="(A)n"

```
repeat_region 9822..9853  
/rpt_family="(CA)n"  
repeat_region 10274..10722  
/rpt_family="MaLR"  
repeat_region 11045..11163  
/rpt_family="CT-rich"  
repeat_region 13136..13287  
/rpt_family="Alu"  
repeat_region 15445..15567  
/rpt_family="AT-rich"  
repeat_region 15850..15875  
/rpt_family="AT-rich"  
repeat_region 15929..16234  
/rpt_family="Alu"  
repeat_region 17161..17194  
/rpt_family="AT-rich"  
repeat_region 18540..18828  
/rpt_family="MaLR"  
repeat_region 18938..19023  
/rpt_family="MaLR"  
repeat_region 19181..19224  
/rpt_family="MaLR"  
repeat_region 19256..19588  
/rpt_family="MaLR"  
repeat_region 19659..19726  
/rpt_family="MaLR"  
repeat_region 19847..20312  
/rpt_family="MER2_type"  
repeat_region 19928..19960  
/rpt_family="(TG)n"  
repeat_region 20313..20623  
/rpt_family="Alu"  
repeat_region 20596..20640  
/rpt_family="(A)n"  
repeat_region 20779..20802  
/rpt_family="AT-rich"  
repeat_region 20868..21156  
/rpt_family="Alu"  
repeat_region 21487..21514  
/rpt_family="AT-rich"  
repeat_region 21958..21983  
/rpt_family="AT-rich"  
repeat_region 22010..22076  
/rpt_family="AT-rich"  
repeat_region 22223..22248  
/rpt_family="(TTTA)n"  
repeat_region 22234..22520  
/rpt_family="Alu"  
repeat_region 23084..23385  
/rpt_family="Alu"  
repeat_region 23503..23691  
/rpt_family="CR1"  
repeat_region 23715..23895  
/rpt_family="MIR"  
repeat_region 24034..24855  
/rpt_family="CR1"  
repeat_region 24966..25190  
/rpt_family="L2"  
repeat_region 25636..25670  
/rpt_family="(CA)n"  
repeat_region 25704..26073  
/rpt_family="MaLR"  
repeat_region 26711..27112  
/rpt_family="MaLR"  
repeat_region 27461..27664  
/rpt_family="L2"  
repeat_region 28756..28814  
/rpt_family="L2"  
repeat_region 29356..29574  
/rpt_family="L1"  
repeat_region 29748..29876  
/rpt_family="ERVL"  
repeat_region 30292..30476
```

```
Query Match 35.0%; Score 808.4; DB 9; Length 165538;  
Best Local Similarity 67.0%; Pred. No. 6.9e-152;  
Matches 1478; Conservative 0; Mismatches 621; Indels 107; Gaps 19;  
/rpt_family="L1"  
QY 194 cggctccgcgggtgggcggcgatcggtcgcgcaaatccgcgctcgccacccggg 253  
DB 49264 CGGTGGCAGCGAGTTGAAGATCGGAGATCAACGTTGAAGATCACCGCTCGCAACCCG 49205  
QY 254 -----cctgcggggcgggcggcgctcgatttctctccctgcctctgcag---cccc 305  
DB 49204 GGCTGGCGCGGGTAGGGCGCGGCGCTCGATTCTCCCTCGCTCCGCCCTCCCT 49145  
QY 306 tgtgcgcatgctcggc-ctacgcggccccagcccttggattga-----tcgggtcggcagcg 359  
DB 49144 GGTGCGCATGCTCAGCTCAGCTCGGCCCTCGCCTTTGATTATTTTCTTGGGGGCC 49085  
QY 360 gctgcgaccctgggcggcgagacggcgggatggggagcc-----cggcgctggg 409  
DB 49084 GCTGGCAGCCGGGACTGACTTCGGGATGGGAAGTGAGAGCCCGGAGCTGCTACCGTGGC 49025  
QY 410 agcggcgagtgatcagc-----ggtggcgccggtgagtcacgggtgagtcacgcgg 461  
DB 49024 GCGCGCGCTGTGAGGAGCAGCCAGGGGAGGAGCTGCGGCTCGCGCTGAGTATCCGGG 48965  
QY 462 -----catggggtccgcaagaagacgcaggaaccccccggtgctgagccacga 512  
DB 48964 AAGCGCCACCATGGGCTCCGTAAGAAGAGACCAAGAACCCCGCTTCTCAGCCAGGA 48905  
QY 513 attcatggtgcagaaccacgcggatggtctcctcgctggtgggcattctctggtgctggg 572  
DB 48904 ATTTCCTCTGCAGAAATCATGCGGACATCGTCTCCGTTGGGATGTTCTTCTCTGCTGGG 48845  
QY 573 acctatgttcgagggcagcgagatgctgatgctgttctcctcaccctcagcagtgagt 632  
DB 48844 TTTTGTGTTTCGAGGGAACAGCAGACATCCATCGTGTGTTCTCACTCTTTCACACAGTGT 48785  
QY 633 cgtgtgcc---cagcggaagggtaccctcggggtccagaccctttaccattatgggt 689  
DB 48784 TGCTGTCCTTCGAGCAGAGGAACAGCCACGGGCTCAAGTCCCTCTATTATTATGTTGT 48725  
QY 690 caaagatctgcccacagtgcttctcatatgctggtggtgcatcatcattcagccaccat 749  
DB 48724 CAAGATTGTGCCACGGTTTCTTCTACATGCTGTTGGCAATCATATTATCATGCCACAAT 48665  
QY 750 ccagagtagctgctagataagctcagcgagactcagctcaccacaaaggaagcaacaaaa 809  
DB 48664 TCAGGAATATGTTGGATAAAATTAACAAGAGAATGAGTTCAACAAAGCGAACAACAAA 48605  
QY 810 caaattgaatgagcgcggtcagctgagtgttctacatagtgctggtggtatctggtggtat 869  
DB 48604 CAAAGTTTAACGAGTCTGGTCAAGTTTAGTGTGTTCTACTTTTTTCTTGTATTGGGGCAC 48545  
QY 870 gatcattcggcctctgagaactcctctcagacccccactctattgtggaagtctcagcc 929  
DB 48544 ATTCAATTTAATCTCTGAAAACCTGCCTCTCAGACCCCAACTCTTATATGAAGGCTCGTCC 48485  
QY 930 ccacaacatgatgacatttcagatgaaattttctacatctcagagttggtggttactgatt 989  
DB 48484 CCATAGCATGATGACATTTCAANTGAAGTTTTTCTACATATATCCAGTTGGCTTACTGGTT 48425  
QY 990 tcatagttcccgagctctactctccagaaagtcaggaacaaagatatcccggtgtaact 1049  
DB 48424 TCATGCTTTTCTGAACTCTACTTCCAGAAAACCAAAAAACAAGACATCCCTCGTCAACT 48365  
QY 1050 catctacatggcctccacctctccacattggagggcctctctcttctgacttgaacaa 1109  
DB 48364 TGCTACATTTGGTCTTCACTCTTCCACATTTACGTGAGGCTTATCTCTTGTACTTGAATCA 48305  
QY 1110 cctgggctgctgctctctgctgctgactatgctgtcgagctcctctccacgctgtgag 1169  
DB 48304 TTTGGGACTTCTCTTTTGGTACTGATATTATTTGTTGAATTACTTTCCACATGTGGGG 48245
```


Qy	792	cacaaaggcaaaacaaacaaattgaatgagccgagcgagtggtgttctacatagt	851
Db	372	CCCCAAGGAAACAAACAAAGTTAAACGAGCTCGTGGTGTAGTGTCTACTTTT	431
Qy	852	gtctgtatctgggtgatgatctattctggcctctgagaactccctgtgcagacccactct	911
Db	432	TTCTTGTTATTTGGGCGACATTCATTTTAATCTCTGAAACTGCCGTGACGCCAACTCT	491
Qy	912	atgttggaagtctagcccccaacaaatgatgacatttcagatgaaatttttctaatctc	971
Db	492	TATATGGAAGGCTCTCCCATAGATGATGACATTTCAATGAAATGAAATTTTCTACATATC	551
Qy	972	acagttggcttactggtttctagtttcccgagctctctacttccagaaagtccagaaaca	1031
Db	552	CCAGTTGGCTTACTGGTTTCACTGCTTTTCTGAACTCTACTTCCAGAAAACCAAAAACA	611
Qy	1032	agataccccgggtcaactcaatcatattggcctccacacctcttccacattggggggccta	1091
Db	612	AGACATCCCTCGTCAACTTGTCTACATTGGTCTTCACTCTTCCACATTACTGGAGCTTA	671
Qy	1092	tctctgtacttgaaacacactgggcctgctgtcttctgtatgctgcactatgctgtcagct	1151
Db	672	TCTCTGTACTTGAATCATCTTGGGACTCTCTCTTGGTACTGCATTAATTTTGTGAATT	731
Qy	1152	cctctccagcgtgtgcagcctcttacttttgggatgagcgtaccagaaaggggtgtc	1211
Db	732	ACTTTCCACATCTGGGCGCTGTTTACTTTAGTGAAGATACCAGAAAGGCATATC	791
Qy	1212	tttggcctatcgtgtttatatccgggagactcgtgacactgattgtctcagtggttac	1271
Db	792	TCTGTGGGCATTTGTTTATCTTGGGTAGACTTGTGACTTTAAATTTGTTCCGTACTCAC	851
Qy	1272	agtagggcttcaacttggccgg--gacaaatcggaatggaaatgctctctcgtgtaagt	1328
Db	852	TGTTGGGTTTCACTTCCCTGGGTGATCGAGAAATCGGAATCTGATGCCCTTACTTGGAAATGT	911
Qy	1328	caatgtgttggcagctaaatcgtcttctcctcagttgcagttcagttatccaggtgtacat	1388
Db	912	AAATGTGTGGCAGCTAAATTTGCTTCTGTCGTCACAGTTGCACGATCCAAAGCTACGT	971
Qy	1389	asactggaccttgacacgcgttgcttcagagatggttagaagatgcgaattcttcagt	1448
Db	972	AAATGGAATCTTAATTACTCTCTGGCTTCAGAGGTGGGTAGAAGATCTTAATATTACGCG	1031
Qy	1449	ctgtgggagaa-----gagacggtccaggtcgagaaagccacagaaatgg	1496
Db	1032	CTCATGTATGAAGAAACGGTCTGAGATCTTCTTAAAAAAGAACACAGAAACGGAGTGGG	1091
Qy	1497	agtgagaaatccaaatagaaatattctccacaaagaagaagaagagctccttagca	1556
Db	1092	AGTGAATACTTCAATAGAGTAGACTGTCCGCCAAAGAGGAAAGAGAAATCTTCATAATC	1151
Qy	1557	gttgcagcgaattgattcttaccctccaaagggaatccacttcttcttattggtgtcct	1616
Db	1152	TTTGAAGCGCATATTAAATGTCTGCAAGGAAT---CTGCTCTTTGAGGTTCTTTCT	1208
Qy	1617	gtgctagaga-tttctgtcttcagaaagcgtcgtgcttttggatttgaattgctcaatgtat	1675
Db	1209	GCACATAGATTTTCTGTTTGTGTTTGAATAATAGTTCGTGCTCTTCGGTTTTGTATTGAAC	1268
Qy	1676	tgctaatgtgttttaa	1693
Db	1269	TGTTTCATGTATTTTAA	1286

RESULT 5
BD005226
LOCUS BD005226 1110 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005226
VERSION BD005226.1 GI:1863187
KEYWORDS JP 03075332-7/1.

SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 1110)		
AUTHORS	Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.		
TITLE	Method of screening of protein		
JOURNAL	Patent: JP 03075332-T 1 16-FEB-2001;		
	SUMITOMO PHARMACEUTICALS CO LTD, NAKOI TODO,HAJIME OKUYAMA, OTOAKI IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO		
COMMENT	OS Homo sapiens. (human)		
	PN	JP 03075332-T/1	
	PD	16-FEB-2001	
	PF	17-AUG-2000	JP 2000005488
	PR	20-AUG-1999	JP 99P 234764
	PI	NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIRONORI ISHIKAWA, KIYOMITSU NEMOTO	
	PC	C12Q1/02.G0IN33/50.C07K14/47.A61K38/17.C12N5/10.C12P21/02// PC (C12P21/02,C12R1:91)	
CC			
PH	Key	Location/Qualifiers	
FT	CDS	(1)..(1107).	
FEATURES	Location/Qualifiers		
source	1..1110		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	296 a	241 c	236 g 337 t
ORIGIN			
	Query Match	28.0%;	Score 647.8; DB 6; Length 1110;
	Best Local Similarity	76.1%;	Pred. No. 4.4e-120;
	Matches 844;	Conservative 0;	Mismatches 247; Indels 18; Gaps 3;
Qy	463 atggggctcgcgaagaagacgcaggaaacccccgggtgctgagcagcaattcatggg	522	
Db	1 ATGGGGCTCGTAAAGAGAGACAAAGACCCCGCTCTCAGCCAGGAATTCATCCTG	60	
Qy	523 cagaaacacgcgcatggtctcctcgtgggcatgttcttcgtcgtgggacttatgttc	582	
Db	61 CAGAATCATGGGACATCGTCTCCTGGTGGGATGTTCTTCTCTGGGCTTGTGTTC	120	
Qy	583 gagggcagcccgagatgctcgatcggttctcactcagcctcagcagatgagtcggtgcc--	640	
Db	121 GAGGGAACACAGAACATCCATCGTGTCTTCTCAGTCTTCTCAGCACAGTGTGTCCT	180	
Qy	641 -cagcgaagggctacccctcgggtccagaccctttaccattatgggttcaaatgactg	699	
Db	181 GCAGCAGAGGAACAGCCAGCGGCTCAAGTCCCTCTATTATTATGTCAGAAATTC	240	
Qy	700 gccacagtggttcttctacatgctggtggccatcatcattccagccaccattcagaggtac	759	
Db	241 GCCACGGTTTTCTTCTACATGCTGGTGGCAATCATTTATTCATGCCACAATTCAG	300	
Qy	760 gtctagataagctcagccgagactgcagctcaccagaaagcaaaacaaatgaat	819	
Db	301 GTGTTGGATATAATTAACAAGAGAATGAGTTCACCAAGAGCAAAACAAAGTTAAC	360	
Qy	820 gagggcggcagctgagtggttctacatagtgctggtatctcgtgggtatgatcatctg	879	
Db	361 GAGTCGGTCAGTGTAGTGTGTCTACTTTTTTCTTGTATTGGGGCACATTCATTTA	420	
Qy	880 gcctctgagaactgcctgtcagacccccactctatttgggaagtcagcccccaaatg	939	
Db	421 ATCTCTGAAACCTGCCTGTGACAGCCCAACTCTTATATGGAAGGCTCGTCCCATAG	480	
Qy	940 atgacatttcagatgaaatttttctacatctcagagttgcttactggttcatggttc	999	
Db	481 ATGACATTTCAATGAAGTTTTCATATATCCAGTGGCTTACTGTTTCTCATGCTTT	540	
Qy	1000 ccggagctctacttccagaaagtcaggaaacagatatcccggtcgaactcatctacatt	1059	
Db	541 CCTGAACCTCTACTTCCAGAAAACCAAAACAGACATCCCTCGTCAACTTGTCTAC	600	

Qy	1060	ggctccacactctccacattgaggggcccctatctcttgtagtgaaccacctggccgtg	1119
Db	601	GGTCTTCACTCTTCCACATTACTACGGAGCTTATCTCTTGTAATGAATCATTTGGGACTT	660
Qy	1120	ctgtctctgatgcgaactatgctgtagagctctctccagcgtgtgcagcctgctttac	1179
Db	661	CTTCTTTTGGTACGTCATTAATTTGTTGAATTAATCTTCCACATGTGCGGCTGTTTTAC	720
Qy	1180	tttgggatgagcgggtaccagaaaggggtgctcttctgtggccctatcgtgttatatccggg	1239
Db	721	TTTAGTGATGAAAGTACCAGAAAGGCATATCTCTGTGGGCCATTTGTTTATCTTGGGT	780
Qy	1240	agactctgacactgattgtctcagtggttacagtagtgaggtcttacttggccgggac---a	1296
Db	781	AGACTTGTGACTTTAAATGTTTCCGTACTACTCTGTTGGGTTTACCTTGGCTGGATCGCAG	840
Qy	1297	aatcggaatgaaatgctctctctctgttgtaagtcaatgtgttggcagctaaatcgctgtt	1356
Db	841	AATCGGAATCTGATGCCCTTACTTGGAAATGTAATGTGTGGCAGCTAAATTTGCTGTT	900
Qy	1357	ctgcctcaggttcagtagtccaggtgtacataacatgagccttgagacogtctgctt	1416
Db	901	CTGTCTCCAGTTCGACGATCCAGCCCTACGTACATGGAATTAATTAATCTCTCTGCTT	960
Qy	1417	cagagatggttagaagatgcgaatcttcctcatctgctgtggaggaa-----gaga	1464
Db	961	CAGAGTGGGTAGAAATTTCTAATATTTCAGGCCCTCATGTATGAAAGAAACGGTCGAGA	1020
Qy	1465	cggctccaggtgcgaaagcagcaaaaatggagtcgagaaatcccaatagaatgattct	1524
Db	1021	TCCTTCTAAAAAGAACAGAAAAACGGAGTGGGAGTGGAAACTTCAATAGAGTACTGT	1080
Qy	1525	ccaccaagaagaagagagaagctccctta	1553
Db	1081	CCGCCAAGAGGAAGAGAAATCTTCATA	1109
RESULT	6		
CFTRAM			
LOCUS	CFTRAM	2856 bp	linear
DEFINITION	C.familiaris mRNA for TRAM-protein.		
ACCESSION	X63678		
VERSION	X63678.1		
KEYWORDS	GI:941		
SOURCE	Tram-protein.		
ORGANISM	dog.		
REFERENCE	Canis familiaris		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
TITLE	Gorlich,D., Hartmann,E., Prehn,S. and Rapoport,T. A. A protein of the endoplasmic reticulum involved early in polypeptide translocation		
JOURNAL	Nature 357 (6373), 47-52 (1992)		
MEDLINE	92244357		
REFERENCE	2 (bases 1 to 2856)		
AUTHORS	Hartmann,E.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f. Molekulare Med., Robert-Roessle-Strasse 10, O-1115 Berlin Buch, FRG		
FEATURES	Location/Qualifiers		
source	1..2856		
	/organism="Canis familiaris"		
	/db_xref="taxon:9615"		
	/cell_line="MDCK"		
	219..1343		
	/codon_start=1		
	/product="TRAM-protein"		
	/protein_id="CAA5217.1"		
	/db_xref="GI:942"		
	/db_xref="SWISS-PROT:Q01685"		
	/translation="MAIRKSTKSPVL SHEFTLQNHADIVSCVMVFLGLMFEITA KASIEFTLQYNTVLPATBEQAEFTSLVYIGIKDLATVFYMLVAIIHAIIOEVITA		
CDS			

BASE COUNT	830 a	517 c	532 g	977 t	
ORIGIN					
Query Match	23.6%	Score 545.4	DB 4	Length 2856	
Best Local Similarity	61.5%	Pred. No. 2.2e-99			
Matches 1183	Conservative	0	Mismatches 661	Indels 81	Gaps 16
Qy	436	ccggtgagtaaccggtgagtaaccgcggaatcggtgggtccgcgaagaagaaagcgcaggaacccc	495		
Db	192	CCTCAGCACAGCGCGCGGCGCTGCCATGCGCATTCGCAAGAAGAGCACCAAGAGCCCC	251		
Qy	496	ccggtgctgagccagcaattcatgctcagaaacacgcggataggtctcctcgctgggc	555		
Db	252	CCGGTCTTGAGCCACGAATCATCTCGAGAAATCAGCGGACATCGTCTCTGTGTGGC	311		
Qy	556	atgtcttcgtctggtggaactatgttcgagggcgcgcgcgagatgctgactggttcctc	615		
Db	312	ATGGCTCTCCGTGGGCTCATGCTCGAGATAACAGCGAAAGCCTCTATCATTTTGT	371		
Qy	616	acctgcagcatgagtggtgtctc --- cagcgaaagggtccacctcggtccaggaacc	672		
Db	372	ACTCTTCAGTACAATGTATACCCTCCCTGCAACAGAAGAACCAAGCTACTTGAATCAACATCC	431		
Qy	673	ctttaccatattgggtcaaagatctggccacagtgtctctcatctgctggtggccatc	732		
Db	432	CTTTATTACTATGGTATCAAGAATTGGCTACAGTTTTCTTCTACATCTAGTGGCGATA	491		
Qy	733	atcattcacgcccacattcaggagtagctgctagataagctcagccggagactcgagctc	792		
Db	492	ATTATTATGATGCATAATTCAAGAGTATGCTGTTGGATAAAATTAACAGCGGAATGCATTTTC	551		
Qy	793	acaaaggccaaacaaacaaatgaatgagggccggcgagctgagtggtctcatcatagtg	852		
Db	552	TCCAAAACGAAACACAGCAAGCTTTAATGAATCTGGTCAAGTCTAGTGCATTTCTACCTTTTT	611		
Qy	853	ctgagtatctggggtatgatcatctcgctctgagaactcctgctcagaccccacacta	912		
Db	612	TC TTGTATTTGGGGCACATTCATTTCTAATATCTGAAAACACTACATCTCAGACCCCAACTATC	671		
Qy	913	tgtggaagtctcagcccccaacatgatgacattccaagatgaatttttcatcatctca	972		
Db	672	TTGTGGAGGGCTTACCCCCATAAACCCTGATGACATTTTCAAATGAAGTTTTCTACATAGCC	731		
Qy	973	cagttgcttactggtttctatagtttcccgagagctctacttccagaaagtcaggaacaa	1032		
Db	732	CAGTTGGCTTACTGGTTCCATGCTTTTCTGTAACCTCTACTTCCAGAAACCAAAAAGAA	791		
Qy	1033	gataccgggtcaactcatcatatggcctccacctcttccacattggaggggcctat	1092		
Db	792	GATATTCCTCTCAGCTGTCTACATGGTCTTTTACCTCTTTCATATGTCGGGGCTTAT	851		
Qy	1093	ctctgttactgaaccacctgggctgctgctctgtatgctgcactatgctgtcgagctc	1152		
Db	852	CTTTTAAACTTGAATCATCTAGGACTTGTCTTCTGGTGGCTACATTAATTTGTTGAATTT	911		
Qy	1153	ctctcagcgtgtgcagcctcttactttggggatgagcggtaaccagaaaggtttgtct	1212		
Db	912	CTCTTCCATATTTCCCGCCTGTTTATTTTATAGTGAATAAGTATCAGAAAGGGTTTTCT	971		
Qy	1213	tgtggcctacgtgtgtttatatccggagagactcgtgacaactgattgtctcagtggttaca	1272		
Db	972	CTGTGGGCAGCTGTGTTGTTTGGGAAGACTTCTGACTTTTAATTTCTTTCAGTACTCACT	1031		
Qy	1273	gtagggtctcacttggccgggaca --- aatcggaatggaaatgctctctctgtgtaatgtc	1329		
Db	1032	GTTGGCTTTGGCCTTGCAGAAGCAGAGAATCAGAAACTTGGATTTTCAGTGTCTGGGAATTC	1091		


```
QY 1330 aatgtgttgagcagtaaaatcgctgtctgtcctcaggtgcagttgcagttacata 1389
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1092 ATGTGTTAGCTGTAGAAATGCTGTCCTGGCATCCATTGTCATCCCAAGCAATCATG 1151
QY 1390 acatgagaccttcagcaccctctggcttcagagatgattagagagatcgaaatcttcgtc 1449
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 ATGTGGAAGTTCATTAATTTTCAGCTTCGGAGGTGGAGGAACATTCCTACTTTTCAGGCA 1211
QY 1450 tttgggaggaag-----agacggtccaggtcgagaaagacagaa 1491
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1212 CCAGTGTGCAAGAAACACACAGTAACATAAGGCAGGTCTCTAGAAAGAACAGAA 1271
QY 1492 aatggagt-----ggagaaatccaaatagaaatattctccacaaagaagaag 1542
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1272 AATGGTGTAAATGGACACAGTAACCTCAATGGAGCTGACTCTCCCGTAATAGAAAGAG 1331
QY 1543 aaagc-tccttagcaggttcagcaggaattgattcttcacctccaaaggaatccactcttc 1601
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1332 AAATCTTCATAATGAATTAATCTAATTAATGAATTAATGTCCCAAGAAATCTGCTTTTA 1391
QY 1602 ttatgtggtctctgtcgtagagatttctctcttcaga--acgggtcggtcttttg 1659
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1392 CTATAT-CTTTCAGCACTAGAGATTTTCTGTCTCTTGCTTGAATACAGTTGTGCTCTTTC 1450
QY 1660 aatattgctaattgtcttaattgtttttaaaggttttcagacagtgatgagtgggga 1719
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1451 ATTTCTTGCTATTGCTGTTGTTTCATGCAAT--TTTTTAAAGGGCATTTGAGGGAAGGA 1508
QY 1720 tgggggttaagactaaacactcagcct-----ctaaatacagtcagaaatag 1766
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1509 TGATTAATCTAGTGAGAGAAAATATTTTAGCTTAGACTAAGCTACCTGCCTTCAAAATAG 1568
QY 1767 ttaacgagcaaacatcttattagttagtctt--acctcaacgattttccaaacgttt 1824
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1569 TTTAGGACCAACACCAATATCTATTTGTTTTTATCTTTGATGATGTTTTTCTTAAGATT 1628
QY 1825 tgtgtgtgact-----gcagaattgtgtacataaataatagtttctcttcccaatgt 1879
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1629 TGGAGAGAAATATTTTACAAAAATCTACATAACAGTGATAAAAATCTTGCTCTCACC 1688
QY 1880 tctttatgaataaacaag-----tctgtcagaaagtggttttttctcactgttct 1933
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1689 AATTTTGTGATTAGCAAGATGCGCTCTCCACGACAGGTCAATTTTAAAGTTATCTTTC 1748
QY 1934 cctgcaggataaagtggaaaactctgataaaggttaaaactcaaatcagttattatgtaacgg 1993
Db 1749 AGGTAAATGGAATACTATACAGTTGGAAGTCAAACTTTAATATGTTTTCAGTGTCT 1808
QY 1994 ttgggatt-----tttttaagtggttttaaatattacaatggaaagcatttgcacaa 2047
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1809 CTAAATTTTAGGAATTTTGTAGTCTTTACACCTGGGAAAAAGATTTGTAATTTCACTA 1868
QY 2048 ccaaaaatattgtttaattttatagtagtagtaattgttagtgcctacgcccccaatt---- 2103
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1869 TAATAATTGCTGCTTTATTTTATAGTAATGTTGTTAATGTTTACATCCCATTTTAAA 1928
QY 2104 --aaagcatcaaaatagaatgacatgtgtgtgtgattgacattttagcgaatcaa 2161
Db 1929 AAACAACCAACACACTATTGTTTACAATGTTGGAGAGTTATTGGCATATATATTGAA 1988
QY 2162 gatacctttaataaatatggtgggttactaaagaagtaaacgacttcttctgtttatt 2221
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1989 TCAAAAATATGAACATAAAG--TACCTGTGAAGTAAAAATCCT--TTCTTGAGCATTT 2044
QY 2222 taacactgtcaggaacactgcgaataataaataattactgataaaaaaataaaaaa 2281
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2045 TCACACTTGTAACTGAAATCAGAAATTAATAATTTTACTATAAAGAACAAATACGA 2104
QY 2282 aaaa 2286
Db 2105 CTATA 2109
```

RESULT 7

```
BC000687
LOCUS
DEFINITION Homo sapiens, translocating chain-associating membrane protein,
clone MGC:784 IMAGE:3347823, mRNA, complete cds.
ACCESSION BC000687
VERSION BC000687.1 GI:12653796
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2722)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc_mgc@nigr.nih.gov
Contact: Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: 1 Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 7657654.
Location/Qualifiers
source
1. 2722
/organism="Homo sapiens"
/db_xref="LocusID:23471"
/db_xref="taxon:9606"
/clone="MGC:784 IMAGE:3347823"
/tissue_type="Kidney, renal cell adenocarcinoma"
/clone_lib="NIH_MGC_14"
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
92. 1216
/codon_start=1
/product="translocating chain-associating membrane
protein"
/protein_id="AAH00687.1"
/db_xref="GI:12653797"
/translation="MATRKSTKSPVLSHEFVLQNHADIVSCVAMVFLGLMFEITA
KASIIIFVTLQYNVTLPAEQATESVLYYGIGKDLATVFYMLVAIIHVAIOEYML
DKINRMHFSKTRKESQJSAFYLFACVWGTFILISENYISDPTILWRAIPHNH
MTQMKFFYISQAWLHAPPELYFQTKKEDIPQLVYIGLYLHFGAGLYLLNHL
GLVLVLHVEFEFLHSRLFYFSEKRYQKFSLWAVLFLGRLTLTLTLVTVGFL
ARAENOKLDEFTGNFNLAVRIAIVASICTQAPMMKFIQFLRRWRHSAFOAPAV
KKKPTVTGSRSSKKTENGVTGLTNSVADSPRNKKEKSS"
```

FEATURES

```
BASE COUNT 834 a 484 c 480 g 924 t
ORIGIN
Query Match 22.9%; Score 529.2; DB 9; Length 2722;
Best Local Similarity 61.7%; Pred. No. 3.9e-96;
Matches 1081; Conservative 0; Mismatches 598; Indels 73; Gaps 12;
```


Qy	436	ccggtgagttaccggtgagttaccgcggtgagttccgcaagaagaacgcccgaagacc	495
Db			
Db	65	CTGTCAGGGCGGGCGGCCCCACCATGGCGATTTCGAAGAAAGACCAAGAGGCC	124
Qy	496	ccggtgctgagccaggaattcaatggtgcagaaacacgcgagataggtctcctcgtggtgc	555
Db			
Db	125	CCAGTCTCTGAGCCACGAATTCGCTCTCGAGAAATCACGGGACATCGTCTCTGTTGGG	184
Qy	556	atgtcttcgtgctgggacctatgttcgagggccgcccagagatgctcgatcgtgttcctc	615
Db			
Db	185	ATGGTCTTCTGCTGGGGCTCATGTTGAGATACGGCAAAAGCTTCTATCATTTTGTGT	244
Qy	616	accctgcagcatggagctgtgtcccgag---cggaaagggtacccctcgggtccagagacc	672
Db			
Db	245	ACTCTTCAGTCAAAATGTCACCTCCCGACCAACAGAAAGACTACTGAATCAGTGTCC	304
Qy	673	ctttacacattatgggtccaaagatctggtccacagtgttctcttcacatgctggtggtccatc	732
Db			
Db	305	CTTTATTTACTATGGCATCAAAATTTGGCTACTGTTTCTTCTATGCTAGTGGGATA	364
Qy	733	atattcacgcaccatcagtagtacctgctagataagctcacgccggaagactgcagctc	792
Db			
Db	365	ATTATTCAATGCCGTAAATCAAAGAGTATATGTTGGATAAAATTAACAGCGCAATGCATTC	424
Qy	793	accaaaggcaacaaacaaatgtaatgagggccggcagctgagctgtgttcttcacatagtg	852
Db			
Db	425	TCCAAACAAACACAGCAAGTTTAATGAATCGGTGAGCTTAGTGGCTTACCTTTT	484
Qy	853	tcgtgtatctgggtgatgatcattctggcctctgagaactgctgtcagaccacactcta	912
Db			
Db	485	GCCTGTGTTGGGCACATTTCTCATCTCTGAAAACTACATCTCAGACCCCAACTATC	544
Qy	913	ttgtggaagtctcagcccccacaaatgatgacatttcagatgaatttttcacatctca	972
Db			
Db	545	TTATGAGGGCTTATCCCATACCTGATGACATTTCAATGAAAGTTTTTCTACATATCA	604
Qy	973	cagttggcttactggtttcatagttcccgagctctctactccagaagtcaggaaacaa	1032
Db			
Db	605	CAGCTGGCTTACTGGCTTCATCGCTTTTCTGNACTCTACTTCCAGAAAACCAAAAAGAA	664
Qy	1033	gataccccgggtgaactcatctacattggcctccacctcttccacattggaggggacctat	1092
Db			
Db	665	GATATCTCCGTGAGCTTGCTTACATTTGCTTTTACCTCTTCCACATTTGCTGGAGCTTAC	724
Qy	1093	ctcttgtaactgaaacacactgggctgctgtcttgatgctgtgcaactatgctgtcagctc	1152
Db			
Db	725	CTTTTGAACCTTGAATCATCTAGGACTTGTCTCTGCTGCTTACATTAATTTGTGTAATTT	784
Qy	1153	ctctcagcggtgcagcctgcttactttgggatgagcggtaccagaaggggtgtct	1212
Db			
Db	785	CTTTTCCACATTTCCGCCCTGTTTTATTTTTAGCAATGAAAGTATCAGAAAGGATTTCT	844
Qy	1213	ttgtggcctatcgtgtttatataccgggagactcgtgacactgattgctcagtggttaca	1272
Db			
Db	845	CTGTGGCGAGTCTTTTGTGTTGGGAAGACTTCTGACTTTTAATTTCTTTCAGTACTGACT	904
Qy	1273	gtagggtctcaacttggccggagc---aaatcggaatggaaatgctctctctgttaatgtc	1329
Db			
Db	905	GTGTGGTTTGGCCCTTCCAGAGACGAGAAATCAGAAAGCTGGGATTTCACTATCGGAAACTTC	964
Qy	1330	aatgtgtggcagctaaatcgtcttctgtcctcgagttgctcagtatccagggtgtacata	1389
Db			
Db	965	ANTGTGTAGCTGTTAGAAATCCGCTGTTCTGGCATCCATTTCCGTTTCTCAGGCAATTTATG	1024
Qy	1390	acatggacctttgacgacctctggctcagagatggttagaagatgcgaatcttctcatgtc	1449
Db			
Db	1025	ATGTGGAAGTTCAATTAATTTTCAGCTTCGAAGTGGAGGGAACATTTCTGCTTTTCAGGCA	1084
Qy	1450	tgtgggaggaagagacggtccaggg-----tcgaaaaaaggaacagaa	1491
Db			
Db	1085	CCAGCTGTGAAGAAGAAACCAACAGTAACATAAAGGCAGATCTTCTCTAAAAAAGGAACAGAA	1144

Qy	1492	aatggagtggag-----aatccaaataagaatagattctctccaccaaagaaagag	1542
Db	1145	AAATGGTGTGAATGGAACAATTAACTTCAAAATGACGACTCTCCCGGAATAAAAAAGAG	1204
Qy	1543	aaagc-tccttagcagttgcaagcgaattgattcttaccctcaagggaatccactcttctc	1601
Db	1205	AAATCTTCATTAATGAATATAAACTAAATGATTAATGTGCCCAAGAAAYCTGTTTCTA	1264
Qy	1602	ttatgtgggtctctgtgctagagatttcttcttcagaaacgggctcgctgtttttgaa	1661
Db	1265	CTATATCTTTCAGCATTAAGAGATTTTCTGTCTGTGAATAACAGCTCTGTCTCTTGAT	1324
Qy	1662	tattgctaattgatgtctaatagtgtttttaaggtttttgcagacgtatgagctgggggatg	1721
Db	1325	TTTTGCTATTGTACGGTTTTCATGCAATTTT-----TTTAAAGGGCATTTGAGGGGAGGATT	1379
Qy	1722	gggttaagaaactaaaccactcagcct-----ctaaatacagtcagaaatagtt	1768
Db	1380	ATTGCTATGAATGAATAAAAAATATTTTACGTTAGACTTAAGCTACCTTGCTTCAAAATAGTT	1439
Qy	1769	aacggaccacaatcttatttagttaggttctta-ccctcaacgattttccaaacgttttgt	1827
Db	1440	TAGGACCACACCATATTTTATTTTGTGTTTATTTTGAACATTTTCTAATGATTTGG	1499
Qy	1828	ggtgatgact-----gcagaattgttacataaaataat--agttctctgcttccaatgtt	1880
Db	1500	AGAGAAAACTATTACAAAAATTCACATATCAGTGATACAAATTTCTTGCCTGTCACCAAT	1559
Qy	1881	ctttatcgaattacaagctcgttagcaaaagtgtgttttctcaatgtttctctcgcag	1940
Db	1560	TTTTTATTAATAGCAGAGTGGCGCTGCTCTTAAGAAAGGCCATATTTTAAAGTTATCTTTCAG	1619
Qy	1941	gataaagtggaaaatc-----tgataaagggttaaaactcaaatcagattatttgt	1988
Db	1620	GGTAACATGGAAATACATATAAAGTTGGATGTCAAACTTTAATATATGTTTCAGTGTCTCT	1679
Qy	1989	aacgcttgggatttttttaaagtgttttaaatttacaatggaaagcattttgtcaaacac	2048
Db	1680	AATTTTGGGAAT-TTTTGTAGACTTTACACCTGGAAAAAAGATTTGTAAAAATCACCGG	1738
Qy	2049	caaaaataigtgttaattttatgtagtagtaattgttttagtcttacgcccccaataaagc	2108
Db	1739	AACAAATCTGTGCTTTATTTTATAGGTAGTGGTATTATAGTATTACATCCCCATTTAAAA	1798
Qy	2109	atcaaaaatatga 2120	
Db	1799	ACAAAAACATAA 1810	
RESULT	8		
AX329795			
LOCUS	AX329795	1267 bp	DNA
DEFINITION	Sequence 304 from Patent WO0194629.		linear
ACCESSION	AX329795		
VERSION	AX329795.1	GI:18102773	
KEYWORDS	.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (sites)		
	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,		
	Horrigan, S., Soppet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature		
	gene sets		
JOURNAL	Patent: WO 0194629-A 304 13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
	Location/Qualifiers		
source	1..1267		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	343 a 278 c 275 g 371 t		
ORIGIN			

Query Match	21.4%;	Score 493.8;	DB 6;	Length 1267;
Best Local Similarity	65.7%;	Pred. No. 4.3e-89;		
Matches 752;	Conservative 0;	Mismatches 387;	Indels 6;	Gaps
Qy	350	gtcggcagcgctgcgaccctggcgcgacacggcggggatgggagcccgcgctggg	409	
Db				
Qy	410	agcggcgagtgatcagcgggtggcgccagccagcggagcggcgagtcggcgagtcg	469	
Db				
Qy	470	tcgcagaagaagcgcaggaaccccccggtgctgagcagcaagaattcattggtgcagaacc	529	
Db				
Qy	530	acgcgatatggtctctcgtcgctgggcatgctctctgctgctgggaactatggttcgagggca	589	
Db				
Qy	590	cggccgagatgtcgatgtgttctctcaacctgcagcatgcgcatgtgctgttgcacg	646	
Db				
Qy	647	aagsgctaccctcggggtccaggaaccttaccattatgggtcacaagatcggccacag	706	
Db				
Qy	707	tgttctctacatgctggtggccatcatctaccgccaccattcaggagtcagtgtctag	766	
Db				
Qy	767	ataagctcacgcggaactgcagctcacaaaggcaaaacaaataaagtgaatgagggcgg	826	
Db				
Qy	827	ggcagctgagctgtgtctacatagctgtcgttatctggggtatgatcattctggcctcg	886	
Db				
Qy	887	agaaactcgtgcagaccccaactctatgtgaagtctcagcccaacaatcatgatgacat	946	
Db				
Qy	947	ttcagatgaaattttctacatctcacagtgtggcttaactggtttcatagtttcccgagc	1006	
Db				
Qy	1007	tcacttccagaagtcaggaaacaagatatccccgggttcaactcatctacattggcctcc	1066	
Db				
Qy	1067	acctctccacatggaggggctactctgttactgaacacacctggcctgtctgttc	1126	
Db				
Qy	1127	tgatgtgcactatgctgtcgagctctctccagcgtgtgcagcctgtttactttgggg	1186	
Db				
Qy	1187	atgagcgttaccagaaagggttcttggcctatcgttttatatccgggagactcg	1246	
Db				
Qy	1247	tgacactgattgtctcagtggtttacagtagggcttcaacttggccgggac---aaatcgga	1303	
Db				
Qy	1304	atggaaatgctctctctggttaatgtcaatgtgttggcagctaaaatcgctgttctgtcct	1363	
Db				
Qy	969	aaactggaatttcagtagtgcgaacttcaaatgtgttagctgtttagaaatcgctgttctgcgat	1028	
Db				

QY	1364	cgaattgcagttaccaggtygtacataaatgaccttgacgccctgtcggcttcagagt	1423
Db	1029	CCATTTGGCTTACTCAGGCATTATGCTGTGAAGTTCATTAAATTTTCAGCTTCGAAGT	1088
QY	1424	qgttagaagatgcgaactctcatgctgtgggagaagacagcggtccagggtcgaaaaag	1483
Db	1089	GGAGGACATTCCTGCTTTTCAGGCACCAGCTGTGAAGAANAACCAACAGTAACATAAG	1148
QY	1484	gcaca 1488	
Db	1149	GCACA 1153	
RESULT	9		
BD005229			
LOCUS	BD005229	1267 bp	linear PAT 31-JAN-2002
DEFINITION	Method of screening of protein.		
ACCESSION	BD005229		
VERSION	BD005229.1	GI:18633190	
KEYWORDS	JP 03075332-T/4.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1267)		
JOURNAL	Todo, N., Okuyama, H., Imamura, M., Ishikawa, H. and Nemoto, K. Method of screening of protein Patent: JP 03075332-T 4 16-FEB-2001; SUMITOMO PHARMACEUTICALS CO LTD, NAOIKI TODO, HAJIME OKUYAMA, OTOAKI IMAMURA, HIROKAWA, KIYOMITSU NEMOTO		
COMMENT	OS Homo sapiens (human) PN JP 03075332-T/4 PD 16-FEB-2001 PF 17-AUG-2000 JP 2000005488 PR 20-AUG-1999 JP 99P 234764 PI NAOIKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIROKAWA, KIYOMITSU NEMOTO PC C12P1/02, G01N33/50, C07K14/47, A61K38/17, C12N5/10, C12P21/02// PC		
FEATURES	source		
BASE COUNT	343 a	275 g	371 t
ORIGIN			

Query Match	21.4%	Score 493.8	DB 6	Length 1267	
Best Local Similarity	65.7%	Pred. No. 4.3e-89			
Matches 752	Conservative	0	Mismatches 387	Indels 6	Gaps 2
QY	350	gtcggcagcggtctgcgaccctggcggcagacggcggggatgggagaccgcgcgtggg	409		
Db					
	9	GGCTGTCAGCGGGCGGTGACACAGCAGCCAGCGGAGCGCGGACGTGCGTGAGCAGCT	68		
QY	410	agcggcgagtgatcagcgggtggcggcgggtgagtagccgtgagtagccgcggcatggggc	469		
Db	69	GGGAAGAGCAGAACCGGGCGGAGCACCTTCAGCGCGGGCGGGCCACCATGCGGA	128		
QY	470	tcgcgaagaagacgcgcaggaaccccccggtgtgtgagccacgaattcatgtgcagaacc	529		
Db	129	TTCCGAAGAAAGACCAAGAGCCCCCAGTGTGTGAGCCACGAATTCGTCTCGAGAATC	188		
QY	530	acgcggatatggtctcgtcgtagggcatgtcttcgtgtggacattatattcgagggca	589		
Db	189	ACCGGACATCGTCTCGTGTGGCGATGCTCTCCGCTGGGGCTCATGTTTGAGATAA	248		
QY	590	cggcccgagatgtcgatggttctctcacctgcgacatggagtcggtgttccccg---	646		

```
Db 249 CGGCAAAAGCTTCTATCATTTTGTACTCTTCAGTAAATGTCAACCTCCAGCAACAG 308
QY 647 aagggtaccctcggggtccagaccctttaccattatgggggtcaagaatctggccaag 706
Db 309 AAGAACAGTACTGAATAGTGCCTTTATTACTATGCGATCAAAAGATTGGCTACTG 368
QY 707 tgttcttctacatgctggtggccatcatcttcacgcccacattccaggagtaogtgcag 766
Db 369 TTTTCTCTACATGCTAGTGGCGATATTATTCATGCGTAAATTCAGAGATATATGTGG 428
QY 767 ataagctcagccgagactcagctaccacaaaggcaacaaacaaatgaatgagccg 826
Db 429 ATAAATTAACAGCGGAATCACTTCTCCAAAACAAAACAGCAAGCTTTAATGAATCTG 488
QY 827 ggcagctgaagtgttctacatagtgctgtatctgtggtggtatgatatcttggccctcg 886
Db 489 GTGAGCTTAGTGGCTGTCTACCTTTTGGCTGTGTGGGCGACATTCATTCTCATCTCTG 548
QY 887 agaactgcctgtcagaccccaactctattgtggaagtctcagcccacacacatgatgacat 946
Db 549 AAAACTACATCTCAGACCCCAACTATCTTATGAGGGGCTTATCCCATAACTGATGACAT 608
QY 947 ttcagatgaatttttctacatctcagctgagcttactgttttcacatagtttcccggagc 1006
Db 609 TTTCAATGAAGTTTCTTACATATACAGCTGGCTTACTGGCTTCACTGCTTTTCTCTGAAC 668
QY 1007 tctacttcagaagaagtcagaaacagatatccgggtcaactcatctacatctacatggcctcc 1066
Db 669 TCTACTTCCAGAAAACCAAAAAGAGATATCTCTGTCAGCTGTCTACATGGGTCTTT 728
QY 1067 acctcttcacattgaggggacctctctctgtacttgaaccacctgggacctgctcttc 1126
Db 729 ACCTCTTCCACATTTGCTGGAGCTTACTCTTTTGAATTTTCCACATTTCCCGCTGTTTATTTAGCA 848
QY 1127 tgaatgcactatgctgtcagctcctcctccagcgtgtgcagcctgttacttcttgggg 1186
Db 789 TGTGTGTACATTTATTTGTTGAATTTCTTTTCCACATTTCCCGCTGTTTATTTAGCA 848
QY 1187 atagcggtaccagaagggttctcttctgtgacctatcggttttatccggagactcg 1246
Db 849 ATGAAAGTATCAGAAGGATTTTCTGTGGCGAGTCTCTTTGTTTGGGAGAGCTTC 908
QY 1247 tgacactgatgtctcagtggtgtacagtaggggttcaacttggccgggac---aaatcgga 1303
Db 909 TGACTTTAAATCTTTCAGTACTGACTGTTGTTTGGCCCTTGGCAAGAGCAGAAAATCAGA 968
QY 1304 atggaatgctctctctgtgtaagtcaatgtgtggcagctaaatacgtctgtctctcct 1363
Db 969 AACTGGATTTTCAGTACTGGAAACTTCAATGTGTAGCTGTAGAAATCGCTGTCTGGCAT 1028
QY 1364 cgaagtgagatccagggtgtacataacatggagacctgacacgctggtcttcagagat 1423
Db 1029 CCATTTGCGTTACTAGGCATTTATGATGGGAAGTTTCATTAATTTTCAGCTTCGAAGGT 1088
QY 1424 ggtagaagatcggaatcttctcatgtctgtgggaggaagagacggtcaggtcagaaaaag 1483
Db 1089 GGAGGGAACATTTGCTTTTCAGGCCACCGCTCTGGAAGAGAAACCAACAGTAACAG 1148
QY 1484 gcaca 1488
Db 1149 GCAGA 1153

RESULT 10
LOCUS HSTRAMP
DEFINITION H.sapiens mRNA for TRAMP protein.
ACCESSION X63679
VERSION X63679.1 GI:37264
KEYWORDS trAm gene; TrAm protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
REFERENCE
AUTHORS Gorlich,D., Hartmann,E., Pehm,S. and Rapoport,T.A.
TITLE A protein of the endoplasmic reticulum involved early in polypeptide translocation
JOURNAL Nature 357 (6373), 47-52 (1992)
MEDLINE 92244357
REFERENCE 2 (bases 1 to 1267)
AUTHORS Hartmann,E.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f. Molekulare Med., Robert-Roessle-Strasse 10, O-1115 Berlin Buch, FRG
FEATURES
source 1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="Hela"
122..1246
/gene="TRAM"
122..1246
/gene="TRAM"
/codon_start=1
/product="TRAM protein"
/protein_id="CAA45218.1"
/db_xref="GI:37265"
/translation="MATRKSTKSPVLSHEFVLQNHADIVSCVANVFLGLLMEFTA
KASLIFVTQYNVTPATERQATESLSLYIGIKDLATVFFYMLVAIIHIAVIOEYML
DTNRRMHFSKTKHSENGSOLFACVVGFTFILISENYISDPTILMRAYPHNL
MKFMKFFYISQALWHLHAFPELYFORTKKEDIPQLVYIGLYFLHAGAYLLNLHL
GLVLLVHLHFEVFLFHLISRLFYFSEKRYQKFSLWAVLVGLRLTLILSLVTGFGFL
ARAENOKLDFSTGNFNVLAIVLASICTVQAPMMKFLNFQLRRRHSFAFOAPV
KKPTVTYTKGRSSKKGTENGYNGLTISNVADSPRNKKESS"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN

Query Match 21.4% Score 493.8; DB 9; Length 1267;
Best Local Similarity 65.7%; Pred. No. 4.3e-89;
Matches 752; Conservative 0; Mismatches 387; Indels 6; Gaps 2;

QY 350 gtcggcagcgctgcgacctgggcgagacgcggggagtgaggagcccgcgctggg 409
Db 9 GGCTGCAGCGGGCGTGCACAGCAGCCAGCGGGAGGCGCGGAGTCCGCTGAGCAGCT 68
QY 410 agcgcgcgagtgatcagcggtgcgcgcggtgagtcacggtgagtcacgcgcgagcggcg 469
Db 69 GGAAGAGCAGAACCGGGCGGAGCAGCTGCAGGCGCGGCGGCGGCCACCATGGCGGA 128
QY 470 tccgcaagaagaacccaggaaccccccggtgctgagccagcaattcatggtgcagaacc 529
Db 129 TTCCCAAGAAAGCACCAGAGCCCCCGAGTCTGAGCCACGAATTCGCTCTGCAGAAATC 188
QY 530 acgcggtatgggtcctcgtcggtggcgatgttctctgctgggaccttatgttcaggcca 589
Db 189 ACGGGACATGCTCTCTGTGTGCGATGCTCTCTCTGCTGGGCTCATGTTTGAGATAA 248
QY 590 cggcgcgagatgcagatcggttctcctaccctgcagcagtcgagtcggtgttccacg---cgg 646
Db 249 CGGCAAAAGCTTCTATCATTTTGTACTCTTCAGTAAATGTCAACCTCCAGCAACAG 308
QY 647 aagggtaccctcggggtccagaccctttaccattatgggggtcaagaatctggccaag 706
Db 309 AAGAACAGTACTGAATTCAGTGTCCCTTTATTTACTATGGCATCAAGATTTGGCTACTG 368
QY 707 tgttcttctacatgctggtggccatcatctaccgcccacctcaggtacgtacgtcgtag 766
Db 369 TTTTCTCTACATGCTAGTGGCGATAAATTTATTCATGCGGTAAATTCAGAGTATATGTGG 428
QY 767 ataagctcagccgagagactcagctaccacaaaggcaacaaacaaatgaatgagccg 826
Db 429 ATAAATTAACAGCGGAATGCACTTCTCCAAAACAAAACAGCAAGTTTAAATGAATCTG 488
```

```
QY 827 ggcagctgagtggtgtctctacatagtgctggtgtatctggtggtatgatcatcttctggtctgtg 886
Db 489 gtcagcttagtcggttttacccttttgcctgtgtgtttggcgccacattctcatctctg 548
QY 887 agaacctgctgtcaagcccccactctattgtgaagctctcagcccccacacaacatgatgacat 946
Db 549 AAACCTACATCTCAGACCCCAACTATCTTATGAGGCGCTTATCCCATACACCTGATGACAT 608
QY 947 ttcatgataaatttttctacatctcacagttggttactggtttctatagtttcccgagac 1006
Db 609 TTCAATGAAGTTTTCTACATATACACCTGGCTTACTGGCTTCATGCTTTCTCTGAC 668
QY 1007 tctactccagaagaatcaggaaacagatatacccggtgcaactcatctacatctggtcctcc 1066
Db 669 TCTACTTCCAGAAAACCAAAAAAGAGATATTCCTCGTCAGCTTGCTACATTTGCTTT 728
QY 1067 acctctccacattgagggcgctctctcttctgtaacttgaaacacacctggcctgctctc 1126
Db 729 ACCTCTTCCACATTTCTGGAGCTTACCTTTTGAACCTTGAATCATCTAGGACTTGTCTTC 788
QY 1127 tgatgtgcaactatgctgcagctcctctccagcgtgtgcagcctgcttactttgggg 1186
Db 789 TGTGCTACATTAATTTTGTGAATTCCTTCCACATTTCCCGCTGTTTATTTTAGCA 848
QY 1187 atgagcgtaccagaaaggttgtcttctgtgacctatgctgtttatatccgggagactcg 1246
Db 849 ATGAAAAGTATCAGAAAGGATTTCTCTGTGGCGAGTCTTTTGTGTTGGGAAGACTTC 908
QY 1247 tgaactgatgtctcagtggttacagtagtgaggtctcacttgccgggac---aaactcga 1303
Db 909 TGACTTTAATCTTCAGTACTGACTGTGTGTTTGGCTTTCACAGAGCAGAAAATCAGA 968
QY 1304 atggaatgctctctctggttaagtcaatgtgttgccagctaaaaatcgctgtctgtct 1363
Db 969 RACTGGATTTCACTAGTAACTGAACTCAATGCTTGTAGCTTTAGATCGCTGTCTGGCAT 1028
QY 1364 cgagttcgagttaccagtggtacataacatgagaccttgacgacctgtggtctcagagat 1423
Db 1029 CCATTTGGCTTACTCAGGCAATTTATGCTGTGAAGTTCATTAATTTTTCAGCTTCGAAGT 1088
QY 1424 ggttagaagatcgaaattctctgtgtgggagagagacggtccaggtcgagaaaaag 1483
Db 1089 GGAGGAACATTCGCTTTTCAGGCACCGAGCTGTGAAGAAAGAAACCAACAGTAACAAAG 1148
QY 1484 gcaca 1488
Db 1149 GCAGA 1153

RESULT 11
BC012401
LOCUS BC012401 2819 bp mRNA linear ROD 20-AUG-2001
DEFINITION Mus musculus, clone MGC:11724 IMAGE:3967323, mRNA, complete cds.
ACCESSION BC012401
VERSION BC012401.1 GI:15214554
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2819)
Direct Submission
Submitted (15-AUG-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC plate: 17 Row: j Column: 12
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

FEATURES

source

Location/Qualifiers

1..2819
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:11724 IMAGE:3967323"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Maml"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
133..1257
/codon_start=1
/product="Unknown (protein for MGC:11724)"
/protein_id="AAH12401.1"
/db_xref="GI:15214555"
/translation="MAIRKSNKNPILLSHEFLONHADIVSCLAMFLLLGLMFEVTA KGAIIFVALQXNVTRPADEQASLYHYGIDKDLATVLFYMLVAIIHAIIOEYVL DKINRMHFSKTKHSESGOLSAYLFCVWGTFILISENYISDPTTLWRAYPHNL MTQKPFYISQLAYWLHAFPELYFKQKKEDIPLQLYIGLYLPHIAGAYLLNLHL GLVLLYHFVEFLHSLRFYSDEKQKGLSWALFVLGLRLTLLILSLVTVFGFL ARKQKLDFTGNFNLAVRIAVLASICTQAFMWKPFINFOLRWRHSHAFQAPPV RRKPATVTKGRSSRKGTENGVTGTSNGADSPRNKRKSS"

BASE COUNT 735 a 597 c 581 g 906 t

Query Match 21.0%; Score 486; DB 10; Length 2819;
Best Local Similarity 66.2%; Pred. No. 1.9e-87;
Matches 752; Conservative 0; Mismatches 370; Indels 14; Gaps 3;

QY 357 gcggctcgacccctggcgccagacggcgggatgggagccgcgcgtggagcgagc 416

Db 35 GCGGAGCAGCCAGCGGAGCGCGCGCGGTGAGCAGCGAGAGCTCGGAGCGGGG 94

QY 417 cagtgatcagcgggtggcgccggtgagttaccggtgagtcacgcggcggtccgcaa 476

Db 95 CGGAGGAAGCACCTGCAGC-----ACAGCGCGCGCTCCACCATGGCGATTGCGAA 146

QY 477 gaagaacgcaggaaccccccccggtgctgagccacgaattcattggtgcagaacacgcgga 536

Db 147 GAAGAGCAACAAGAACCCCGCGCTGCTGAGCACCAGCAATTTCTTGTCTCAGAAATCACGCGGA 206

QY 537 tatgtctcctcgtggcgatcttctcgtctgggacctatgttcgagggcagcgccga 596

Db 207 CATCGTCTCTCGCTGGCGATGCTTTTCTCTCTGGGGCTTATGTCGAGGTACAGCAAA 266

QY 597 gatgtcagtcgtgttctcacccctcagcatggagtcggttctcc---cagcggaaggcgt 653

Db 267 AGGAGCCATCATCTTTGTGCTCTTCAGTATAATGTTACCCGCCCTGCAACAGAACA 326

QY 654 accctcggggtccagacccttaacattatgggtcagaagatctggtccacagtgctct 713

Db 327 AGCTACTGAATCAGCATCCCTATATACATATGCTATCAAAAGACTTTGGCTACAGTTTGT 386

QY 714 ctacatgctggtggccatcatcattcacccaccattcagagtagctagatgaagct 773

Db 387 CTACATGCTAGTGGCGATAATTTATTCATGCCATAATTCAGGAGTATGTGTGGATAAAT 446

QY 774 cagccggagactcagctcaccagggcaacaaacaaatgaatgagccggcgagct 833
 Db 447 TRACCGACGGATGCACTTCTCCAAACGAGACACGAAGTTTAAACGAGTCTGGTCAGCT 506
 QY 834 gagtggtttacatagtgctgtggtatctctggtggtatgatattctggtcctctgagaactg 893
 Db 507 CAGTGCATCTACCTTTTGGCTGTGTGTGGGCACAGTTTCATTCTCATCTCCGAAACCTA 566
 QY 894 cctgtcagaaacactctatttggaagctcagcccaacacatgatgacatttcagat 953
 Db 567 TATCTCAGACCCCAACATCTCTGTGGAGAGCTATCCCTCGTCACTCGTCTACATTTGCGCTTTATCTCTT 626
 QY 954 gaaattttctacatctcacagcttggtcttactggtttcatagtttcccgagctcactt 1013
 Db 627 GAAGTTTTCTACATCTCCAGCTGGCGTACTGGCTCCCATGATTCCTTGAACCTCTACTT 686
 QY 1014 ccagaaagtccaggaacacagatccccgggtccaaactcactcactcactcactcctt 1073
 Db 687 CCAGAAAACCAAAAAAGAGATATCCCTCGTCACTCGTCTACATTTGGCCTTTATCTCTT 746
 QY 1074 ccacattgaggggctctctcttcttacttgaaacacacctggcgctgctctctgatgct 1133
 Db 747 TCAATATGCTGGAGCCTATCTCTGAACCTGGAACCACTGGGCTTGTCTCTTCTGTGCT 806
 QY 1134 gcactatgctgcagctcctcctccagcgtgtgcagcgtgcttactttggtgggatgagcg 1193
 Db 807 GCACATTTTCGTTGAATTTCTTTTCCACATTTCTCGTCATTTTATTTAGTCACGAGAA 866
 QY 1194 gtaccgaaagggtgtcttctgtggtcctatcggtgtttatatccgggagacgtctgacact 1253
 Db 867 GTATCAGAAAGGGTTTCTCTCTGGCAGTCTCTTTTGTGTTTGGGAAGACTTCTGACTCT 926
 QY 1254 gattgtcagtggttacagtaggcttcacttggc---cggaacaaatcggaatggaaa 1310
 Db 927 AATCTTTTCACTTCTTACGTGTGGTTTGGTCTCGCAAGACCGAGAAATCAGAAAGCTGGA 986
 QY 1311 tgcctctctggaatgtcaatgtgtgtggcagctaaatcgctgttctgctccctcagttg 1370
 Db 987 CTTTACTTGGAAACTTCAACGCTTGGCGGTTAGGATGCGTGTCTCGCGTCCATTG 1046
 QY 1371 cagtatccgggtgtacataacatggacctgacgacctgtgcttcttcagagatggttaga 1430
 Db 1047 CATCACACAGCCCTCATGATGTGGAAGTTTCATTAACTTCCAGCTTCGGAGGTGGAGGGA 1106
 QY 1431 aaatgcgaattctcgtctgtgtggggaagacagcaggtccaggtccagaaagga 1486
 Db 1107 ACATTCTGCTTCCAGGCCCCACCTGTGAAAAGGAAACCGGCCGTGACCAAGGCA 1162

RESULT 12

BTU19578 2387 bp mRNA linear MAM 06-SEP-2000
 LOCUS Bos taurus translocation chain-membrane associating protein (tram)
 DEFINITION mRNA, partial cds.
 ACCESSION U19578
 VERSION U19578.1 GI:9971727
 KEYWORDS
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 2387)
 Clark, T.G., Morris, J., Akamatsu, M., McGraw, R.A. and Ivarie, R.D.
 Cloning and sequence analysis of a bovine tram cDNA
 Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 2387)
 Clark, T.G., Morris, J., Akamatsu, M., McGraw, R.A. and Ivarie, R.D.
 Direct Submission
 JOURNAL Submitted (06-JAN-1995) Genetics, University of Georgia, Athens, GA
 30602-7223, USA
 FEATURES
 Location/Qualifiers
 1..2387 source

/organism="Bos taurus"
 /db_xref="taxon:9913"
 <1..2387
 /gene="tram"
 <1..1077
 /gene="tram"
 /codon_start=1
 /product="translocation chain-membrane associating protein"
 /protein_id="AAG10391.1"
 /db_xref="GI:9971728"
 /translation="EFVLQNHADIVSCVMVFLLGLMEITAKVSIIFVTLQYNVTLPL
 ATEQATSEAFLYYGIKDLATVFFVLMVAIIHAIQEVYLDKINRMHFSKTKHAK
 FNSGOLSAFLYFCISLWGTILISENVISDPTILRAYPHNLMTFOKMFYISOLAYW
 FHAPFLYFQKTKKEDIPOLVYIGLYLHAGAYLLNLNLGLVLLVLAHVFVLELH
 ISRLVFSDEKYQKGSRLWAVLVLGRLLLLIISLYTVGSLARAENQKLDFTSGNEN
 VLARITAVLASICITQAFMMKIKFQLRWRHSAFQAPAVKKKPPVTKGRSXXKGT
 ENGVRTVTSNGADSPNRKEKS"
 BASE COUNT 675 a 473 c 449 g 788 t 2 others
 ORIGIN

Query Match 20.8%; Score 481.6; DB 4; Length 2387;
 Best Local Similarity 64.0%; Pred. No. 1.4e-86;
 Matches 807; Conservative 0; Mismatches 419; Indels 34; Gaps 4;
 QY 511 gaattcattggtcagaacacacgcgagatggtctcctcgtggtgagcattcttcgtgctg 570
 Db 1 GAATTCTGCTGCAAGATCAGCGGACATGCTCTCTGCTGTGGCGATGCTCTCTGCTG 60
 QY 571 ggacttatgttcgagggcagcgagatgtcgatcggttctcctcacccctcagcatgga 630
 Db 61 GGGCTCATGTCGAGATAACAGCAAAAGTTTCTATCATTTTGTACTCTTCAGTACAAT 120
 QY 631 gtctgtgtccag---cggaagggtaccctcgtgggtccagagaccttaccattatgg 687
 Db 121 GTTACCCTCCCTGCGACAGAGAACAAGCTACTGAATCAGCGTTCCTTTATTACTATGT 180
 QY 688 gtcaaatgctggcacagtgcttctacatgctgtgtggccatcatcatcagccacc 747
 Db 181 ATCAAAGATTGGCTACAGTTTCTCTACATGCTAGTGGCGATANTTATTATGCCATA 240
 QY 748 attcaggagtacgtgctagataaagctcagccgagactgcagctcaccaaaaggcaacaa 807
 Db 241 ATTCAAGATATGTTGGATAAATTAACAGCGGAATGCATTTCTCCAAAACGAAACAC 300
 QY 808 acaaaattgaaatgagccggcagctgagtggtgtctacatagtgctcgttatctggggt 867
 Db 301 AGCAAGTTTAAATGAATCTGGTTCAGCTTAGTGCATTTACCTGTCTTGTATTGTTGGC 360
 QY 868 atgatacttgcctctgagaactgctgtcagacccacctctatgtggaagctcag 927
 Db 361 ACATTCAATCTATPATCTGAAAACCTACATCTCAGACCCGACTATTTGTGGAGGCTTAT 420
 QY 928 cccacaacatgatgacatttcagatgaaaattttctacatctcacagttggttactgg 987
 Db 421 CCCATAACTGATGACATTTTCAGATGAAGTTTTTCTACATATACACAGTGGCATACTGG 480
 QY 988 ttctaatgttcccgagctctactctcagaagtcaggaagcaagataatccccgggtcaa 1047
 Db 481 TTTTCATGCTTTTCTCGAATCTACTTCCAGAAAACCAAAAAGAGATATTCTCTCGTCAG 540
 QY 1048 ctatctacattggcctccacctcttccacaattggaggggctctatctcttacttgaac 1107
 Db 541 CTCGCTACATCGGCTTTACCTCTTTCACATTTCTGAGAGCTTACCTTTTAAACTTGAAT 600
 QY 1108 cactggcgctgctctctgactgactgactgctgctcagctccctctcctcagcgtgctg 1167
 Db 601 CATCTAGAGCTTGGCTTTTGGTGTCTACATTTTGTTCGAGATTCTTTTCCACATTTCC 660
 QY 1168 agcctgttctactttggggatgagcgttaccagaagggttcttctgtggtcctatcgtg 1227
 Db 661 CGCCTATTTTACTTTACTGACGAAAAGTATCAGAAAGGATTTTCTCTGTGGCAGTCTG 720


```
Db 969 ATGTGGAGTTTCATTAACCTTCCAGCTTGGAGGTGGAGGGAACATCTTGCCTTCCAGGCC 1028
Qy 1450 tgtgggaggaagacagcgtccaggtcagagaaagca 1486
Db 1029 CCACCTGTGAAGGAACCGCGGTGACCAAGGCA 1065

RESULT 14
AY029534 1338 bp mRNA linear VRT 02-MAY-2001
LOCUS Xenopus laevis Traml mRNA, complete cds.
DEFINITION AY029534
ACCESSION AY029534
VERSION AY029534.1 GI:13936286
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1338)
Hartmann, E.
Direct Submission
Submitted (10-APR-2001) Biologie, University Lubeck, Ratzeburger
Allee 160, Lubeck 23538, Germany
Location/Qualifiers
FEATURES
source 1..1338
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="RZPD A0180202"
16..1137
/codon_start=1
/product="Traml"
/protein_id="AAK40302.1"
/db_xref="GI:13936287"
/translation="MGIRKKNKTPPVLSHEFTIQNHADIVSLCMVFLGLMFETIA
KVAVFVTLQYNTIPVEGLVPTSLYHYGIKDMATVFFYMLVAIILHVAIOEYVLD
KINRMHFSKTHSKNESQLSAFYFSCIGWASIIVSSENFSDPISLWKGYPHYVF
PFQKFFYISOLAVPHARPELFYOKTKEDIPOLVYIGLYLFIHGIAGVAVLNLRLG
LVLLVLYFVEFLFHMRSRLFYSNERYKQFTVWAVLFVLGRLLILLSLVLTGFGFLA
RAENQELDSNGFNFIILIRITICITQAFMMKFIINFQLRNREHSSPOFPSSOR
KKATSAKGASRKREKENVGTITSNGDSPRSRKEKHS"
BASE COUNT 362 a 321 c 267 g 388 t
ORIGIN

Query Match 19.08; Score 438.6; DB 5; Length 1338;
Best local Similarity 63.28; Pred. No. 5.6e-78;
Matches 691; Conservative 0; Mismatches 399; Indels 3; Gaps 1;

Qy 462 catgggctcgcgaagaacgcaggaaccccccggtcgtgagccacgaattcatggt 521
Db 15 CATGGGCATCGTAAGAAGAACACAGACCCCGCGGTGCTACGCCAGGATTCATAT 74

Qy 522 gcagaaccacgcgcatatggtctcctcgtggtggcattcttctcgtggtgggactatgtt 581
Db 75 CCAGAACCATGCGGACATGCTCTCTCGCTGGCCATGCTCTCTCGCTCGGCCCTATGTT 134

Qy 582 cgaaggcacggccagatgtcgatcgtgttctcaccctgcagcagcagcgtcgtgtgcc 641
Db 135 CGAGATTACAGCAAGGTGGTGTGATGTTGTGCACATCCAGTCACAAATGTCAACATTC 194

Qy 642 agcggaaggctaccctcggggtcagaccctttaccattatgggtcgaagatcggc 701
Db 195 AGTAGAAGGGGTGTTGGAGAACCCACTCGCTGACATTTATGGCATCAAGATATGGC 254

Qy 702 cacagtgtcttctacatgctggtggccatcatcattcaaccgccaccatcaggatcagt 761
Db 255 CACTGTATTTTACATGCTGGGGCAATAATACTACTAGTCAGTTCATCCAGAAATAGCT 314

Qy 762 gctagataagctcagcggagactgcagctcaccacaaaggcaacaaacaaatgaatga 821
Db 315 ATTAGATAAAATCAACCGCGGTATGCATTTTTCACAGCAAAACACAGCAAGTTTAATGA 374
```

```
Qy 822 ggcggggcagctgagctgtgttctacatagctgtctgtgtatctgtgggtgatcatcttgcg 881
Db 375 GTCGTGACAGCTAAGTCGATTTTACCTGTTCTCTTGGCATTTTGGGGAGCAAGCATATTCT 434
Qy 882 ctctgagaactgcctgctcagacccacctctatttggaaagtctcagcccccacaaatgat 941
Db 435 CTCAGAGAATTATTTTTCGGATCCAATAAGCCTGTGGAAAGGGTATCCTTCACACATACT 494
Qy 942 gacattcagatgaatttttctacatctcacagttggcttactggttttctagtttcc 1001
Db 495 TCCATTCAGATGAAGTCTTCTACATTTCCAGCTGGCTTATTTGGTTTCATGCTTTTCC 554
Qy 1002 ggagctctacttccagaagtcaggaacaagatatccgggtcgaactcatcatctg 1061
Db 555 AGAGCTCTACTTTCAGAAAAACGAAAGAGACATTCACAGGAGCTGGTCTACATTGG 614
Qy 1062 cctcacctcttccacatggaggggcctatctcttctgtacttgaaccacacggcctgct 1121
Db 615 TTTATACCTCTTCCATATTATAGAGCTTATGTTCTCAACCTCAATCGCTGGCTGGT 674
Qy 1122 gctctgactgctgactatgctgcagctcctcctccagcgtgagcgtcttcttactt 1181
Db 675 ACTGCTAGTCTGCAATTACTTTGTAGAAATCTCTGTTCATATGTCAGCTCTGTTTACT 734
Qy 1182 tgggatgagcgggtaccagaaagggtgtgtcttctgtggcctatcgtgtttatatccgggag 1241
Db 735 CAGCAATCAGAGATACACAGAAAGGGTTTACTGTGTGGCGCTTCTCTTTGCTCTTGGGG 794
Qy 1242 actcgtgacactgattgtctcagtggttacagtagaggttccacttggcgggac---aaa 1298
Db 795 CTTCTCTCACTCTTATCTCTTCTGTTCTTACTCTCGCTTTGGACTTGCACGGCTGAAAA 854
Qy 1299 tcggaatggaaatgctctctcgttaagtcaatgtgttggcagctaaatcgctgttct 1358
Db 855 TCAAGAGTTGGCATCGAATAATGGAATTTCAACATTTTGGCCATCAGAATCACCGTGT 914
Qy 1359 gtccctcgagttgcagatccaggtgtacataacatggaccttgagaccgtctggcttca 1418
Db 915 GGCATCTATTTGCATCACCAGGCATTTATGATGTGGAAGTTTCATTAATTTCCAGCTGAG 974
Qy 1419 gagatggttagaagatcgcaatcttcctcgtctgtgtgggaggaagacggtccaggtc 1478
Db 975 ACGATGGAGAGAGCATTCATCTCCTCAACCCCTCCTCTCAAGGGAAGAGCGGACCTCTGC 1034
Qy 1479 aaaggccacagaaaaatggagtgagagaatcccaatagatatagattctccaccacaaagaa 1538
Db 1035 TAAAGGCAAGCTTCACCTAAAGAAAAAGAAATGGAGTGAATGGAACAGTAAACATCAAA 1094
Qy 1539 agagaagctcct 1551
Db 1095 CGAGCAGATTTCT 1107

RESULT 15
AY029529 1453 bp mRNA linear VRT 02-MAY-2001
LOCUS Danio rerio Traml mRNA, complete cds.
DEFINITION Danio rerio
ACCESSION AY029529
VERSION AY029529.1 GI:13936276
KEYWORDS zebrafish.
SOURCE Danio rerio
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 1453)
Hartmann, E.
Direct Submission
Submitted (09-APR-2001) Biologie, University Lubeck, Ratzeburger
Allee 160, Lubeck 23538, Germany
Location/Qualifiers
FEATURES
source 1..1453
/organism="Danio rerio"
```



```
/db_xref="taxon:7955"
/clone="MPMGp609N2214Q8"
79..1188
/codon_start=1
/product="nram1"
/protein_id="AAK40297.1"
/db_xref="GI:13936277"
/translation="MGIRKTKNPPVLSHEFVIONHADVSCVAMVFLGLMFETTS
KVAVLFITVQYNTIPANGPEEATVNYFHYGLKDVATFFYMLVAIIHAIIOEVL
DKINKMHSEKTKHSEKFNESGOLSAFYLFESCLWGASILGSENILSNPVSJWDGYPHYL
MQFQMKFYICQGLGFWLHAPELIFQKAKKEDIIPROLVIISILYLHVAIGALNLRLL
GVLVLVHYFLHFIHSLRIYFSNEERQSGTVMVAVFLVGLRLTLISLVITVGFGL
AGAERQGLNLAEGSNVLFVRVTYLAATCITQAFMWKFIINFQLRRWRQAOQTQLKK
KSSSSKSKSKKANGVSGVSGASGDSPRARKEKS"
BASE COUNT      371 a   340 c   323 g   419 t
ORIGIN

Query Match      16.9%; Score 390.8; DB 5; Length 1453;
Best Local Similarity 62.7%; Pred. No. 2.3e-68;
Matches 643; Conservative 0; Mismatches 377; Indels 6; Gaps 2;

QY 463 atggggtccgcaagaagacgcaggaacccccgggtgctgagccacgaattcatgtg 522
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 ATGGGCATCCGAAAAAGACCACCAAGAACCCCGCGGTGCTCAGCCATGAGTTTGATC 138
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 523 cagaacacgcggtatggtctctcggtggcatgttcttctgctggtgacttatgttc 582
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 CAGAACACGCGGATATTTGTGCTCTGTGCGTATGGTGTCTCTGCGTCTTAATGTTT 198
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 583 gagggcacgcgcagatgtcgatgtgttctcaccctgcagcagtgagtcgtgtccca 642
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 GAGATTACATCAAAAGGTAGCGGTTTGTTCATAAATGTGCAATACAAACGFCACGATTCCA 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 643 gcggaaaggctaccctcggggtccaggaaccccttac---cattatgggtcacaagatctg 699
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 GCAAAATGGAGGTCCAGAGGAAGGCACAGTGAATTACTTTCAATTATGGTCTGAAAGATGTG 318
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 700 gccacagtgctctctacatgctgggtggccatcatcatcacccaccattcagggagtc 759
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 GCCACCATTCTTTTACATGTTGGTGGCCATTATCATGCACGCCCATTTATTCAGGAATAT 378
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 760 gtgctagataagctcagcgggagactgcagctccacaaaggcaaaaacaaaatitgaat 819
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 GTGCTGGATAAAATCAACAGGAAGATGCACTTCTCTAAAAACCAAGCACAGCAAGTTCAAT 438
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 820 gaggcgggcagctgagtggttctacatagtgctggttatctggtgggtatgatcattctg 879
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 GAGTCGGGTGAGTGAGTGCCTTCTACCTCTTCTCTGCTGTGGGGAGCCAGCATTCCT 498
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 880 gcctctgagaactgctgcagaccaccaactctattgtggaagtctcagccccaacaacatg 939
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 GGTCTGAGAAACATCTCTCAAAACCTGTGAGCCTGTGGGATGGTTACCCACACACCCTA 558
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 940 atgacatttcagatgaattttctacatctcacagtttggttacttggttttcatagtttc 999
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 559 ATGCAGTTCCAGATGAAGTTCTACTACATCTGTCACTCGGCTACTGGCTTTCATGCAATC 618
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1000 ccggagctctacttcagaaaagtcagaaacaagatatccccgggtccaactcatctacatt 1059
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 619 CCAGAGCTTTACTTCCAGAAAGGCCAAGAAAGACATTTCTCGCCAGCTTGTTTATATA 678
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1060 ggcctccacctctccacattgaggggacctatcttcttacttgaaccacctgggacctg 1119
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 679 AGCCTTTACTTGGTCCACATFCGCTGAGCCTACGCTCTGAATCTGAACCGGTTGGGTTTG 738
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1120 ctgcttctgatgctgcactatgctgtcgagctctctccagcggtgtgcagcctgctttac 1179
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 739 GTGTTGCTGGTCTGCTGATTTCTTGTGAGCTCATCTTCCATGTGTCAGACTCATTTAC 798
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 1180 tttggggatgagcgggtaccagaaaagggttctgttttggcctatctgttttatatccggg 1239
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 799 TTCAGCAATGAGGAAAGGCAGAGTGGGTTTACAGTGTGGGCTGCTTCTGTGTTGCTGGGC 858
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 1240 agactgtagacactgattgtctcagtggttacagtaggggtctcacttggccgggaca--- 1296
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 859 CGGCTGTTGACTCTGTCTCTGTCTTCTTACTGTGTGGCTTCGGCCTTGACAGGACGAGAG 918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1297 aatcggaatggaaatgctctctctgtgtaatgtcaatgtgttgagcagctaaaaatcgctgt 1356
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 919 AAACAGGGTTTGAATCTTGCAGAGGGAAGCTTTAATGTCCTGTTGTGCGAGTGACCGTT 978
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1357 ctgtctctcagattgcagtatcccaaggtgtacataacatggaccttgacgacccgtctgtgctt 1416
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 979 CTGGCTGCCATATGCATCAAGCTTTTATGATGTGGAATTTATCAACTTCCAGTTG 1038
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1417 cagagatggttagaagatgcgaaatcttctcatgtctgtgggaggaagacggtccagggtcg 1476
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1039 CGACGATGGAGAGAACACAGCACAGACTCAGACCTTAAAGAGAATAATCATCTCTTCTAAG 1098
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1477 agaaaa 1482
||| |||
DB 1099 AGCAAA 1104
```

Search completed: September 6, 2002, 14:46:34
Job time: 36398 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 12:13:06 ; Search time 495.6 Seconds
(without alignments)
8006.032 Million cell updates/sec

Title: US-09-807-470-1
Perfect score: 2311
Sequence: 1 agagagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2311	100.0	2311	21 AAA38012	Rat WAR-1 nucleoti
2	1092	47.3	1092	22 AAF74781	Rat WAR-1 encoding
3	802.2	34.7	2288	21 AAA38013	Human WAR-1 nucleo
4	770.6	33.3	1835	23 AAS76421	DNA encoding novel
5	718.2	31.1	1736	22 AAK94181	Human full-length
6	695.2	30.1	1288	22 AAF74782	Human WAR-1 encodi
7	647.8	28.0	1110	22 AAF74780	Human WAR-1 encodi
8	532	23.0	2833	21 AAC77810	Human cancer assoc
9	517.8	22.4	2756	24 AAS62269	CDNA sequence #56

10	477.6	20.7	1517	24 AAS18576	cDNA encoding huma
11	414	17.9	1120	20 AAZ33531	Human prostate can
12	372.2	16.1	802	22 AAK91674	Human CDNA 5'-end
13	372.2	16.1	802	22 AAK93837	Human CDNA clone r
c 14	345	14.9	642	22 AAS47389	Human breast cance
c 15	345	14.9	642	22 AAF17959	Human breast cance
16	260.6	11.3	1380	22 AAL01678	Human reproductive
17	212.6	9.2	1526	23 ABL18595	Drosophila melanog
18	212.6	9.2	1549	23 ABL18061	Drosophila melanog
19	212.6	9.2	2153	23 ABL21096	Drosophila melanog
20	212.6	9.2	3346	23 ABL21098	Drosophila melanog
21	212.6	9.2	3618	23 ABL18060	Drosophila melanog
22	212.6	9.2	4146	23 ABL18594	Drosophila melanog
23	212.6	9.2	4775	23 ABL21091	Drosophila melanog
24	212.6	9.2	5055	23 ABL21094	Drosophila melanog
25	212.6	9.2	11287	23 ABL21090	Drosophila melanog
26	202	8.7	472	21 AAC00195	Human secreted pro
c 27	109.4	4.7	545	22 AAK92492	Human CDNA 3'-end
c 28	78	3.4	271	22 AAL23061	Human breast cance
c 29	77.2	3.3	422	22 AAL15924	Human breast cance
30	76.6	3.3	272	22 AAL14199	Human breast cance
31	69.2	3.0	402	22 AAI64805	Membrane-bound pro
32	68.2	3.0	3265	21 AAZ65095	Human DNA encoding
33	68.2	3.0	3265	22 AAS46053	Human PRO1124 cDNA
34	68.2	3.0	3265	22 AAF92092	Human PRO1124 (UNQ
35	68.2	3.0	3265	22 AAF44241	Human secreted pro
36	68	2.9	1119	21 AAC60033	Human immune syste
c 37	67.8	2.9	18011	24 ABL32034	Human CDNA encodin
38	67.6	2.9	1254	22 AAS25886	Human breast cance
c 39	67.4	2.9	852	22 AAL19845	Human breast cance
c 40	67	2.9	545	22 AAH70126	Human cervical can
41	67	2.9	958	22 AAC92233	Apoeaquin-encodi
~42	67	2.9	958	24 AAD22188	Aequorea victoria
43	67	2.9	959	9 AAN81534	PAQ440 aequorin ge
c 44	66.8	2.9	310	22 AAH71505	Human cervical can
c 45	66.6	2.9	1952	21 AAC59577	Human secreted pro

ALIGNMENTS

RESULT 1
AAA38012
ID AAA38012 standard; DNA; 2311 BP.
XX
AC AAA38012;
XX
XX 22-AUG-2000 (first entry)
DT
DE Rat WAR-1 nucleotide sequence.
XX
XX Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; rat; ds.
XX
XX Rattus norvegicus.
OS
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
XX
XX WPI; 2000-317980/27.
DR P-PSDB; AAY98146.
XX
PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell

Db 1127 ttgggaactctctcttctgtgtagctgattatttggtagaattacttccacatgtagcgg 1186
Qy 1170 cctgctttactttgggtagcagcgtaccagaaagggttcttcttggcctcattggtt 1229
Db 1187 cctggtttacttttagtgaaagaccagaaaggcatatctctgtagcctcattggtt 1246
Qy 1230 tatatccgggagactcgtgacactgattgtctcagtggttacagtagggcttcaacttgc 1289
Db 1247 tatctgggtagacttgactttaattgtttccgtactcactgttgggtttcacctgcg 1306
Qy 1290 cgg---gacaaatcgaatggaatgctctctctcgtgtaattgtaattggtgagcgttaa 1346
Db 1307 tggatcgcagaatcggaatcctgatgcccttactggaatgtaattgtttggcagcgttaa 1366
Qy 1347 aatcctctctctcgtcagtgatccagtgatccaggtgtacataacatggaaccttgacgac 1406
Db 1367 aattgctgtctcgtccagctgcacgatccagcctcaacatgtaacatggaacttaattac 1426
Qy 1407 cgtctggctcagatggttagaagatgcgaatctctcatgtctgtgaggagaa----- 1460
Db 1427 tctctggctcagagtggttagaattcttaattcgaagcctcatgtatgaaaaagaa 1486
Qy 1461 -----gagacggtccaggtcgagaaaggcacagaaatggagtggagaatccaaatag 1514
Db 1487 acggtcgagatcttctaaaaaaagacagaaacggagtgggagtggaacttcaaatag 1546
Qy 1515 aatagattctccacaaagaaagaaagaaagcctccttagcaggttgcaagcgaattgatt 1574
Db 1547 agtagactgcgcgaagagaaagaaatcttcataatctttgcaagcgcattgatt 1606
Qy 1575 cttaacctcaagggaatccactctctctatgtgtgtctctgtctgtctagaga-ttttctg 1633
Db 1607 aatgctcgaaggaat---ctgctcttgaggtttcttctgctagagattttctcg 1663
Qy 1634 tcttcagaacgggtcgtgcttttgaatattgctaattgatttctaattggtttt--- 1691
Db 1664 ttttgaaaatagtcgtgctctcggtttttgttattgaaactgtttcatttcttca 1723
Qy 1692 -----aaggtttgcagacgtatgagtgggggtatgggggttaagacta 1734
Db 1724 aagacattggaggagagattattatgaatgggaaaaaagattttggttgagacta 1783
Qy 1735 aacctcagctctaaatcacagtcagagaatagtttaacgg---accaactcttatttag 1790
Db 1784 aattactcgtcaaaaataatgcaaaatagttttggggtacccactactatttggtt 1843
Qy 1791 ttgagttcttaacctcaacgattttccaaacgttttgggtgtagctgcagaaattgtgta 1850
Db 1844 tgatttttaaccttcaacattttcttaattgattgtagagataactgcacaattttgca 1903
Qy 1851 cataaataatagt-----ttcgtgctccaatggttcttttatgtaattacaagtcgtcta 1905
Db 1904 tataatgactgctgttcttaactccacagctgtttctcaataactaaacagatggtctct 1963
Qy 1906 gcaagtggttctttctcaatgttctcgtcagagataaagtgaataatcgtataaagg 1965
Db 1964 cctgcaagattatgtgttaatgctctgttcttgggggttaaa---taaaagtcagaaaagg 2021
Qy 1966 ttaactcaaatcagttattgtaaccgtttggatttttttaaaagtttttaaaattaca 2025
Db 2022 tggaaagtaaatcagttctgttaattgttagaattatttttttaagaacttacaactca 2081
Qy 2026 atgaaagcattttgcacaacaccca-aaatatgtgtttaaatttttatgtagtagtaattgt 2084
Db 2082 gaaaagattgttagactaccacaaaataaataatgttctttatttttacaggtagtgattat 2141
Qy 2085 tagtgcctacccccattaaagcatcaaaataatgaatagatgacatgtgtggtgatat- 2143
Db 2142 tagtgccttcaccccatttaaaaaaacacagctactaa-tgggttaacacatatggaggtttg 2201
Qy 2144 ---tgacatttagcgaa-tcaagataaccttttaataatat 2179

Db 2202 ctgccatatattgcatcaaaatatcatattaataat 2240

RESULT 4
AAS76421
ID AAS76421 standard; cDNA; 1835 BP.
AC AAS76421;
XX
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12225.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR P-PSDB; ABG12234.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 1: SEQ ID No 12225; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1835 BP; 533 A; 343 C; 356 G; 603 T; 0 other;
SQ

Query Match 33.3%; Score 770.6; DB 23; Length 1835;
Best Local Similarity 68.4%; Pred. No. 4.9e-140;
Matches 1268; Conservative 0; Mismatches 519; Indels 66; Gaps 12;

Qy 463 atgggggctccgcaagagacgcaggaacccccgggtgctgagccacgattcaggtg 522
|||||
Db 1 atggggctcgtgaagagcaccagaccccccggttctcagcaggaattcactg 60

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y; Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H; WPI; 2001-524255/58.
P-PSDB: AAM93265.

830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation -

Claim 8; SEQ ID NO 2724; 1380pp + sequence listing; English.

The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.

Sequence 1736 BP; 448 A; 358 C; 390 G; 540 T; 0 other;

Query Match 31.1%; Score 718.2; DB 22; Length 1736;
Best Local Similarity 72.2%; Pred. No. 6.9e-130;
Matches 1052; Conservative 0; Mismatches 363; Indels 46; Gaps 8;

QY	435	gcgggtgagtaacggtgagtacccggtcagtcacggtcgatggggctccgcaagaagaacgccaggaaacc	494
Db	157	gcgggtgagtattccggga-agcgccacctaggggtccgtgaagaagacccaagaacccc	215
QY	495	cccggtgctgagccacgaattcatggtgcagaacacacgcggaatatggttctcctgcgtggg	554
Db	216	ccccgttcagccaggaaattcatcctcagaaatcatcgcgacatcgtctcctgcgtggg	275
QY	555	cattgtttcttgctgctgggacttatgttcgagggcacgcccagagatgctgatcgttgtct	614
Db	276	gatgtttcttcctgctgggctgtgttcgagggaacagacaagaatcatcgtgtttct	335
QY	615	cacctgcagcatygagtgctgtgtccc---cagcgaaagggtacocctcgggggtccaggac	671
Db	336	cactctcagcacagttgtctgtccctgcagcagaggaacaagccacggtctcaaagtc	395
QY	672	cccttacattatgggttcaaatgctggcacagatctggcacagatgtctcttcatactgctgggtgccaat	731
Db	396	ccctctattatgggttcaagaatttggtccacggtctctcttcatactgcctgggtggcaat	455
QY	732	cattcttcagccacattccaggagtagcgtgctagataagctcagcggagagactgcagct	791
Db	456	cattattcatgccaeattcaggaaatggtgtgataaaaattaacagaagaaatgcagtt	515
QY	792	caccaaaagcacaacaaaaaaattgaatgagccgggcagctgagtggttctacatagt	851
Db	516	cacciaagcgaacaaaaaacagtttaaagcagtcgtggtcagtttagtgttctactttt	575
QY	852	gtctggtatctgggtgatgatcattctggccctctgagaactgcctgtcagaccacctct	911
Db	576	tctctgatttggggcaactcattttaatctctgaaactgcctgtcagaccacctct	635
QY	912	attgtggaagttctcagcccaacatgatgacatttcagatgaaatttttctacatctc	971
Db	636	tatatggaagctcgtcccatagcatgatgacatttcacaaatgaagtttttctacgtatc	695
QY	972	aacgttggcttactggtttctatagtttcccgagggtctctacttcagaaagtccagaaaca	1031
Db	696	ccagttggtctactggtttcatatgttttcttgaaactctacttcagaaaaacccaataaca	755


```

ID AAF74780 standard; cDNA; 1110 BP.
XX AC
XX AAF74780;
XX
DT 17-MAY-2001 (first entry)
XX DE
XX XX
XX Human WAR-1 encoding cDNA sequence SEQ ID NO:3.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder; ss.
XX OS
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1110
FT /tag= a
FT /product= "WAR-1"
XX
XX WO200114582-A1.
XX
XX 01-MAR-2001.
XX
XX 17-AUG-2000; 2000WO-JP05488.
XX
XX 20-AUG-1999; 99JP-0234764.
XX
XX (SUMU ) SUMITOMO PHARM CO LTD.
XX
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
XX WPI: 2001-202940/20.
XX P-PSDB: AAB70695.
XX
XX Transformation of a cell with separate vectors expressing the sense and
XX antisense strands of WAR-1 DNA for screening secretory and membrane
XX proteins expressed by the cell -
XX
XX Claim 2: Page 62-65; 79pp; Japanese.
XX
XX The present invention describes a screening method for secretory and
XX membrane proteins consisting of transformation of a cell with separate
XX expression vectors for the sense and antisense RNA of DNA encoding an
XX endoplasmic reticulum membrane protein participating in endoplasmic
XX reticulum transport of proteins. Also described are: (1) secretory and
XX cell membrane proteins identified by the screening method; (2) drug
XX compositions containing these proteins; (3) host cells transformed by
XX the separate expression vectors of the method; and (4) the preparation
XX of secretory and cell membrane proteins by culture of the transformants.
XX The method can be used for the identification and preparation of
XX proteins for use in the treatment and prevention of diseases such as
XX cancer, disorders of the nervous system, immune disorders (including
XX allergies and rheumatism) and skeletal disorders. The present sequence
XX encodes a specifically claimed human WAR-1 protein from the present
XX invention.
XX
XX Sequence 1110 BP; 296 A; 241 C; 236 G; 337 T; 0 other;
XX
XX Query Match 28.0%; Score 647.8; DB 22; Length 1110;
XX Best Local Similarity 76.1%; Pred. No. 2.7e-116;
XX Matches 844; Conservative 0; Mismatches 247; Indels 18; Gaps 3;
XX
XX 463 atgggggtccgcaagaagcgcaggaaccccccggtgtgtgagccagcaatcattggtg 522
XX 1 atgggggtccgtaagaagagcaccagaaccccccggtgtgtcaccggaattcattcctg 60
XX
XX 523 cagaacacacggataggtctcctggtgggcaatgttcttcgtgggaactatgttc 582
XX 61 cagaatcgcgacatcgtctcctcgtggggatgttcttctcgtcgtgggctgtgttc 120
XX
AAAF74780 standard; cDNA; 1110 BP.
XX AC
XX AAF74780;
XX
DT 17-MAY-2001 (first entry)
XX DE
XX XX
XX Human WAR-1 encoding cDNA sequence SEQ ID NO:3.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder; ss.
XX OS
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1110
FT /tag= a
FT /product= "WAR-1"
XX
XX WO200114582-A1.
XX
XX 01-MAR-2001.
XX
XX 17-AUG-2000; 2000WO-JP05488.
XX
XX 20-AUG-1999; 99JP-0234764.
XX
XX (SUMU ) SUMITOMO PHARM CO LTD.
XX
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
XX WPI: 2001-202940/20.
XX P-PSDB: AAB70695.
XX
XX Transformation of a cell with separate vectors expressing the sense and
XX antisense strands of WAR-1 DNA for screening secretory and membrane
XX proteins expressed by the cell -
XX
XX Claim 2: Page 62-65; 79pp; Japanese.
XX
XX The present invention describes a screening method for secretory and
XX membrane proteins consisting of transformation of a cell with separate
XX expression vectors for the sense and antisense RNA of DNA encoding an
XX endoplasmic reticulum membrane protein participating in endoplasmic
XX reticulum transport of proteins. Also described are: (1) secretory and
XX cell membrane proteins identified by the screening method; (2) drug
XX compositions containing these proteins; (3) host cells transformed by
XX the separate expression vectors of the method; and (4) the preparation
XX of secretory and cell membrane proteins by culture of the transformants.
XX The method can be used for the identification and preparation of
XX proteins for use in the treatment and prevention of diseases such as
XX cancer, disorders of the nervous system, immune disorders (including
XX allergies and rheumatism) and skeletal disorders. The present sequence
XX encodes a specifically claimed human WAR-1 protein from the present
XX invention.
XX
XX Sequence 1110 BP; 296 A; 241 C; 236 G; 337 T; 0 other;
XX
XX Query Match 28.0%; Score 647.8; DB 22; Length 1110;
XX Best Local Similarity 76.1%; Pred. No. 2.7e-116;
XX Matches 844; Conservative 0; Mismatches 247; Indels 18; Gaps 3;
XX
XX 463 atgggggtccgcaagaagcgcaggaaccccccggtgtgtgagccagcaatcattggtg 522
XX 1 atgggggtccgtaagaagagcaccagaaccccccggtgtgtcaccggaattcattcctg 60
XX
XX 523 cagaacacacggataggtctcctggtgggcaatgttcttcgtgggaactatgttc 582
XX 61 cagaatcgcgacatcgtctcctcgtggggatgttcttctcgtcgtgggctgtgttc 120
XX
QY 583 gagggcagggccgagatgtcgatcgtgttctccacccctcagcatgagtcgtgttcc-- 640
DB 121 gagggaaacagcagaagcatccatcgtgttctcactcttcagcacagtgctgtccct 180
QY 641 -cagcggaaagggtaccctcgggtccagcactttacattatgggtgcaagatctg 699
DB 181 gcagcagaggaacacagccagcggctcaagtcctctattattatggtgcaagattg 240
QY 700 gccacagtggtttctacatgtgtgtggccatcatcattcagccaccattcaggatc 759
DB 241 gccacggtttttctacatgtgtgtggcaatcattattcagccacaaattcaggat 300
QY 760 gtgctagataagctcagccgagactcagctcaccacaaaggcaacaaacaaatgaat 819
DB 301 gtgttgataaaataacaaagagaatgcagttcaccacaaagcgaacaaacaaatgaat 360
QY 820 gagggccggcagctgagtggtttctacatagtggtgttatctgggtgatgattctg 879
DB 361 gagtctggtcagtttagtggttctactttttctgtatttggggcacattcatttta 420
QY 880 gctctgagaactgctgtcagaccccaactctatttgggaagtctcagccccaacatg 939
DB 421 atctctgaaactgctgtcagaccccaactctatttgggaagtctcagccccaacatg 480
QY 940 atgacatttcagatgaataattttctacatctcacagttgcttactggtttctatg 999
DB 481 atgacatttcgaatgaatgttttctacatctccacagttggttactggtttctatg 540
QY 1000 ccggagctctactccagaagtcagggaacaaagatatccccgggtcacaactcatcat 1059
DB 541 cctgactctactccagaacaaacaaacaaacaaacaaacaaacaaacaaacaaacaa 600
QY 1060 ggcctccacctctccacattggaggggcctctctcttacttgaaacacacctggcctg 1119
DB 601 ggtctccacctctccacattggaggggcctctctcttacttgaaacacacctggcctg 660
QY 1120 ctgctctgactgctcactatgctcagctcctcctcagcgtgtgagcctgcttacc 1179
DB 661 ctctcttggtactgactattttgtgaattacttccacacatgtgcggcctgtttac 720
QY 1180 ttggggatgagcgtaccagaaggtgtgttctgttggtcctatgctgtttatccggg 1239
DB 721 tttagtgatgaaagtaccagaagcattctctgtggggcattgtgtttctgtgtt 780
QY 1240 agactcgtgacactgattgtctcagtggtgtacagtagggtcacttgccgggac--a 1296
DB 781 agactgtgacttgaattgttccgtactcactgttgggttccacctgggtggtcag 840
QY 1297 aatcggaatggaatgctctctctgtgtaagtcaatgtgttgagcagtaaaatcgtgt 1356
DB 841 aatcggaatcctgagccttacttggaatgtaaatgtgttgagcagtaaaatcgtgt 900
QY 1357 ctgtcctcagttgagtcattccaggtgtgtacataacatggaccttgacgacctgtgt 1416
DB 901 ctgtcctcagttgagtcattccaggtgtgtacataacatggaccttgacgacctgtgt 960
QY 1417 cagagatggttagaagatgcgaattcttctgtgtggaggaa-----gaga 1464
DB 961 cagagtggtgagaagattcctaattcaggcctcatgtatgaaagaacaggtcgaga 1020
QY 1465 cgttcaggtcgagaaagccagaaatggagtgagagatcccaataagatagattct 1524
DB 1021 tctctaaaaaagaaacagaaacagagtgaggagtggaacttcaaatagagtagactgt 1080
QY 1525 ccaccaagaagaagaagaagactcctta 1553
DB 1081 ccgccaagaagaagaagaagactcctta 1109
XX
RESULT 8
AAC77810
ID AAC77810 standard; cDNA; 2833 BP.
XX
```


Dbb 571 aaaacatcatctcagaccocaaactattttatgaggaggttatccccataacacctgatgaat 630
Qy 947 ttcagatgaatattttctacatctcacagttggtctactggtttcatagttttcccgagc 1006
Dbb 631 ttcaaatgaagtattttctacatcatcacagctggtctactggtttcatgcttttctctgaac 690
Qy 1007 tctacttcagaagaatcaggaaacagatataccccgggtcaactcatctacattggcctcc 1066
Dbb 691 tctacttcagaaaacccaaaagaagatattccctgcgtcagctgtctacattggtcttt 750
Qy 1067 acctttccacattgaggggacctatctcttgaactgaacacacctggcctgctgcttc 1126
Dbb 751 acctttccacattgctgagctgacttcttgaactgaacacacctggaactgctgcttc 810
Qy 1127 tgatgctcaactatgctgagctctctccagctgctgcagctgctgcagctgctttactttgggg 1186
Dbb 811 tgggtctacattattttggtgaattctttccacatttcccgctgttttatttttagca 870
Qy 1187 atgagcgtaccagaaaaggtttgtttgtggtccctatggtttatatccgggagactcg 1246
Dbb 871 atgaaaagtatcagaaggattttctctgtggcgagttttttgttttgggaagacttc 930
Qy 1247 tga-cactgattgtctagtggttacagtagggttcacttggccgggac---aaatcgg 1302
Dbb 931 tgaaccttaattcttctcagtagctgtgtgttgggcttgcagagcagaaaatcag 990
Qy 1303 aatgaaaatgctctctctgtgtaattgtcaatgtgttggtgagcttaaaatcgctgtctgtcc 1362
Dbb 991 aagctggtatttcagtaactggaacttcaatgtgttagctgttagaatcgctgtttctggca 1050
Qy 1363 tggagttcagtatccaggtgtacatacatggaacctgtgacacgctgtggtctcagaga 1422
Dbb 1051 tccatttgcgttactcaggcatttatgatgtggaagttcatttaatttccagcttcgaaag 1110
Qy 1423 tgggttagagatgcgaacttctatctgtctgtgggaggaagagagcgtccagg----- 1473
Dbb 1111 tggagggaacatttctgttttcaggccaccgctgtgagaagaacccacagataactaaa 1170
Qy 1474 -----tcgagaaaagccacagaaaatggagtgag-----aatccaaataga 1515
Dbb 1171 ggcagatcttctaaaaaggacagaaaatggtggaatgggaacatttaactcacaatga 1230
Qy 1516 atagatttccaccaaaagaaaagagaaagc-tccttagcagttgcaagcgaattgatt 1574
Dbb 1231 gcagactctcccggaataaaaaagagaaatcttcataatgaattataaaactaattgatt 1290
Qy 1575 cttaactccaagggaatccacttcttcttattgtgtctgtctgtctgtctagagattttctgt 1634
Dbb 1291 aatgtccccaagaatactgcttctactatcttccagcattagagagatttttctgttc 1350
Qy 1635 tcttcagaaacgggtgctgtcttttgaattattgcttaattgttcttaattgtttttaag 1694
Dbb 1351 ttgaaaaacagctgtgctgtctgttttattgttttattgtacggttttcagctttt--- 1406
Qy 1695 gttttgcagacgtatgagtgggggttgggttagactaaaccactcagcct----- 1747
Dbb 1407 -ttttaaaggcattgaggggaggtattgtctatgaatgaaatgaaaaaatatttttagcttag 1465
Qy 1748 -----ctaaatcacagtcagaatagtttaacggacacacatcttatttagttggttcta 1801
Dbb 1466 actaagctaccgtccctcaaaaatagtttagggaccaccacacatttttattttgttttta 1525
Qy 1802 cctca-acgattttccaaacgttttgggtgatgactg-----cagaattgtgtacataa 1855
Dbb 1526 tttttgaacatttttctaatgatttggagagaaaactatttcaaaaatttccacatatca 1585
Qy 1856 ataataag---tttctgcttccaaatgttctttatcgaaattaaacagctgctgctgcaagtg 1913
Dbb 1586 gtgacaaatttctgtctcaccattttttataatgacagagtggtgctgttcttaagaa 1645
Qy 1914 gttgttttctcaatgttctctcagagataaagggtgaaatc-----tgata 1961
Dbb 1646 ggcataatttttaagtttatcttccagggttaacatgggaataactataaagttggatgtca 1705

Qy 1962 aaggtttaaactcaaatcagctatttatgtaaccgttgggatttttttaaagtgttttaaat 2021
Dbb 1706 aactttaaatgttttcaagtggtctctaatatttttggaaattttgt-agactttacacct 1764
Qy 2022 tacaatggaagcatttgcataaccaccccaaaaatatgtgttattttatagtagtaaat 2081
Dbb 1765 ggaaaaaaagatttgaataccaccggaacaattgtgtgtttatttataggtagtggt 1824
Qy 2082 tgttaagtgttaagcccccattaaagcatcaaaataga 2120
Dbb 1825 tattagattatcatccccatttttaaaacaaaacataa 1863
RESULT 10
AAS18576
ID AAS18576 standard; cDNA; 1517 BP.
XX
AC AAS18576;
XX AC
DT 12-MAR-2002 (first entry)
XX DE cDNA encoding human translocating chain-associated membrane protein.
XX KW Human; translocating chain-associated membrane protein; BioTRAM; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 153..790
FT FT /*tag= a
FT FT /product= "Translocating chain-associated membrane
XX PN CN1310184-A.
XX PD 29-AUG-2001.
XX PF 24-FEB-2000; 2000CN-0111729.
XX PR 24-FEB-2000; 2000CN-0111729.
XX PA (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.
XX PI Mao Y, Xie Y;
XX WPI: 2002-034947/05.
XX P-PSDB; AAU10977.
XX PT New human transposition chain related membrane protein and its coding
XX PT sequence -
XX PS Claim 6; Page 18; 22pp; Chinese.
XX CC The invention relates to a novel human translocating chain associating
XX CC membrane protein (BioTRAM), polynucleotides encoding this polypeptide
XX CC and the recombination process used to produce the polypeptide. The
XX CC present invention also discloses the method of applying the polypeptide
XX CC and polynucleotides in treating immunological disorder, malignant tumour,
XX CC cancer and other diseases. The antagonist resisting the polypeptide and
XX CC its treatment effect is also disclosed. Diagnosis and determination
XX CC method based on the discrimination of the mutation in the nucleic acid
XX CC sequence and the change in the polypeptide expression level, and the
XX CC application of the polynucleotides encoding the BioTRAM. The present
XX CC sequence represents the coding sequence of the novel human BioTRAM
XX CC protein as described in the invention.
XX SQ Sequence 1517 BP; 459 A; 262 C; 308 G; 488 T; 0 other;

Query Match 20.7%; Score 477.6; DB 24; Length 1517;
Best Local Similarity 64.6%; Pred. No. 2.7e-83;
Matches 913; Conservative 0; Mismatches 439; Indels 62; Gaps 11;

Query Match		14.98;	Score 345;	DB 22;	Length 642;
Best Local Similarity		72.0%;	Pred. No. 1.le-57;		
Matches 450;		Conservative 0;	Mismatches 175;	Indels 0;	Gaps 0;
Qy	641	cagcgggaagggctaccctcgggggtccaggaccctttaccattatgggggtcaaaagatctgg	700		
Db	625	CAACAGAAGACACAGCTACTGAATCAGTGTCCCTTTATTACTATGGCATCAAGATTGG	566		
Qy	701	ccacagtgttcttatacatgctggtggccatcatcaccgcccaccattcaggaggtacg	760		
Db	565	CTACTGTTTTCTTCTACATGCTAGTGGCGATAATTTATTCATGCCGTAATTCGAAGAGTATA	506		
Qy	761	tgctagataagctcagccggagagctcagctcaccaaaggcaaaacaaataatgaatg	820		
Db	505	TGTTGGATAAAATTAACAGGCGAATGCATCTCTCCAAACAAAACACAGCAAGTTTAATG	446		
Qy	821	aggcggggcagctgagtggtttcacatagtgctggttatctggggtatgatcattctgg	880		
Db	445	AATCTGGTCAGCTTAGTGGCTTACCTTTTGGCTGTGTTGGGGCACATTCATCTCA	386		
Qy	881	cctctgagaactgcctctcagacccccactctattgtggaagtctcagccccacaacatga	940		
Db	385	TCTCTGAAACTACATCTCAGACCCAACTATCTTATGAGGGCTTATCCCATTAACCTGA	326		
Qy	941	tgacatttcagatgaataattttctacatctcacagttggttactggttttcattagtttc	1000		
Db	325	TGACATTTCAAATGAAGTTTTTCTACATATCACAGCTGGCTTACTGGCTTCATGCTTTTC	266		
Qy	1001	cggagctctactctcagaagaagtcaggaacaagatatccccgggtcaactcatctacattg	1060		
Db	265	CTGAACCTACTTCCAGAAACCAAAAGAGATATTCCTCGTCAGCTTGTCTACATTG	206		
Qy	1061	gcctccacctctccacattggaggggccctatctcttgaattgaaccacctgggcccgc	1120		
Db	205	GTCCTTACCTCTCCACATTGCTGGAGCTTACCTTTTGAACCTGAATCATCTAGGACTTG	146		
Qy	1121	tgctctgatgctgactatgctcagctccctcctccagcgtggtgagcctgctttact	1180		
Db	145	TTCTTCTGGTGCTACATTAATTTTGTGAATTTCTTTCCACATTTCCCGCCTGTTTAAAT	86		
Qy	1181	ttggggatgagcgggtaccagaaaaggttggctcttggcccttatcggttttatccggga	1240		
Db	85	TTAGCAATGAAAGTATCAGAAAGGATTTTCTCTGTGGGCAGTTCTTTTGTGTTGGGAA	26		
Qy	1241	gactcgtgacactgattgtctcagt	1265		
Db	25	GACFTCTGACTTAAATCTTTTCACT	1		

Search completed: September 6, 2002, 14:57:08
Job time: 9842 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 13:08:16 : Search time 97.09 Seconds
(without alignments)
5846.730 Million cell updates/sec

Title: US-09-807-470-1

Perfect score: 2311

Sequence: 1 agagagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	2.9	958	2	US-08-757-046A-5
2	67	2.9	958	3	US-09-447-208-5
3	67	2.9	958	3	US-09-135-988-5
4	67	2.9	958	4	US-09-277-716-5
5	67	2.9	958	4	US-08-597-274A-5
6	64.2	2.8	1474	4	US-08-821-994-64
7	64	2.8	1582	3	US-08-545-196B-10
8	64	2.8	1582	3	US-08-545-196B-12
9	63.4	2.7	2674	4	US-09-817-180-1
10	63.4	2.7	2852	3	US-09-027-137-2
11	63.2	2.7	1441	4	US-08-821-994-63
12	63	2.7	1315	4	US-09-721-822A-10
13	63	2.7	1512	2	US-08-909-965C-8
14	63	2.7	5173	1	US-08-242-677-1
15	62.8	2.7	2447	2	US-09-014-969-14
16	61.4	2.7	1454	4	US-09-372-422A-19
17	60.8	2.6	1882	4	US-09-370-253-1
18	60.6	2.6	2862	1	US-08-148-209A-1
19	60.6	2.6	2989	6	5378464-1
20	60.4	2.6	1364	1	US-08-265-087-3
21	60.4	2.6	1364	1	US-08-621-493-3
22	60.4	2.6	1364	2	US-08-965-688-3
23	60.4	2.6	1364	4	US-09-260-173-3
24	60.2	2.6	144	1	US-08-702-344-26
25	60.2	2.6	240	1	US-08-628-417-6
26	60.2	2.6	1066	1	US-08-157-101A-4
27	60.2	2.6	1736	3	US-09-182-816-22

c 28 60.2 2.6 1736 3 US-09-182-816-24 Sequence 24, Appl
29 60.2 2.6 1736 3 US-09-471-528-22 Sequence 22, Appl
c 30 60.2 2.6 1736 3 US-09-471-528-24 Sequence 24, Appl
31 60.2 2.6 1736 4 US-09-634-530-22 Sequence 22, Appl
c 32 60.2 2.6 1736 4 US-09-634-530-24 Sequence 24, Appl
33 60.2 2.6 2407 4 US-09-370-807-7 Sequence 7, Appl
34 60 2.6 536 1 US-08-341-568-1 Sequence 1, Appl
35 60 2.6 536 2 US-08-911-020-1 Sequence 1, Appl
36 60 2.6 1493 1 US-08-340-820-24 Sequence 24, Appl
37 60 2.6 1493 1 US-08-593-535-24 Sequence 24, Appl
38 60 2.6 3080 4 US-09-099-041A-25 Sequence 25, Appl
39 59.8 2.6 2671 6 5168051-9 Patent No. 5168051
40 59.6 2.6 1172 1 US-07-945-288-9 Sequence 9, Appl
41 59.6 2.6 1172 1 US-08-462-831-9 Sequence 9, Appl
42 59.6 2.6 1172 1 US-08-461-809-9 Sequence 9, Appl
43 59.6 2.6 1172 1 US-08-461-441-9 Sequence 9, Appl
44 59.6 2.6 1172 5 PCT-US93-08518-9 Sequence 9, Appl
45 59.6 2.6 1198 3 US-09-248-335-27 Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-757-046A-5
: Sequence 5, Application US/08757046A
: Patent No. 5876995
: GENERAL INFORMATION:
: APPLICANT: Bryan, Bruce
: TITLE OF INVENTION: BIOLUMINESCENT ARTICLES OF MANUFACTURE
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Brown, Martin, Haller & McClain
: STREET: 1660 Union Street
: CITY: San Diego
: STATE: CA
: COUNTRY: USA
: ZIP: 92101-2926
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/757,046A
: FILING DATE: 11-25-96
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/597,274
: FILING DATE: 02-06-96
: ATTORNEY/AGENT INFORMATION:
: NAME: Seidman, Stephanie L
: REGISTRATION NUMBER: 33,779
: REFERENCE/DOCKET NUMBER: 6680-105B
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 619-238-0999
: TELEFAX: 619-238-0062
: TELEX:
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: FEATURE:
: NAME/KEY: Coding Sequence
: LOCATION: 115..702
: OTHER INFORMATION: apoaequorin-encoding gene


```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 115...702
; OTHER INFORMATION: apoaequorin-encoding gene
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: 5,093,240
; AUTHORS: Inouye et al.
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 82
; PAGES: 3154-3158
; DATE: (1985)
US-08-597-274A-5

Query Match      2.9%; Score 67; DB 4; Length 958;
Best Local Similarity 63.2%; Pred. No. 4.6e-07;
Matches 103; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2149 tttagcgcaatcaagataaccttaataatattggtgggtactaaagaagtaaacgacttc 2208
Db 792 TTGAACGATTCAATCGTTTGTGTTGATTGTTGTAATTAGGAACAGATTAATAATCGAATGA 851

Qy 2209 ttctgtttatttaaacactgttacaggaaactcgcaaaattaaattactgaaaaa 2268
Db 852 TTAGTGTGTTTTTAATCAACAGAACTTACAAATCGAANAAGTAAAAAAGTAAAAAAGTAAAAA 911

Qy 2269 aaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
Db 912 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 954
```

```

RESULT 6
US-08-821-994-64
; Sequence 64, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
; APPLICANT: Greenland, Andrew J
; APPLICANT: Thomas, Didier RP
; APPLICANT: Jepson, Ian
; TITLE OF INVENTION: Promoters
; FILE REFERENCE: PPD 50108
; CURRENT APPLICATION NUMBER: US/08/821,994A
; CURRENT FILING DATE: 1997-03-22
; EARLIER APPLICATION NUMBER: PCT/GB97/00729
; EARLIER FILING DATE: 1997-03-18
; EARLIER APPLICATION NUMBER: GB 9606062.9
; EARLIER FILING DATE: 1996-03-22
; NUMBER OF SEQ ID NOS: 89
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 64
; LENGTH: 1474
; TYPE: DNA
; ORGANISM: Brassica napus
US-08-821-994-64
```

```

Query Match      2.8%; Score 64.2; DB 4; Length 1474;
Best Local Similarity 49.0%; Pred. No. 2.8e-06;
Matches 171; Conservative 0; Mismatches 178; Indels 0; Gaps 0;

Qy 1963 aggttaactcaaatcagttattgtaacccgttggttttttaaaagtttttaaaattt 2022
Db 1116 aagtacagtccaaacccgtttcgtggaaccagcagttgttgaagtttttaaaataaaact 1175

Qy 2023 acaatgaaagcatttgtcaaacaccacaaataatgtgtttaattttatagtagtaatt 2082
Db 1176 caattgggagttttataactaaagatttaattctcatattattgtttgtatgtagtatat 1235
```

```

Qy 2083 gttagtcttacgccccattaaagcatcaaaaatgatgaatagatgacatgtgtgtgata 2142
Db 1236 caaaaaagaaggtatttgatccagcatcaggtatttagctataaatcccttatgtcgat 1295

Qy 2143 ttgacatttagcgaatcaagataacctttaataatattggtgggtactaaagaagtaaac 2202
Db 1296 caatatcatttcgttcaaaagaagatgattgtgtttattgtatttaagaagaagtataa 1355

Qy 2203 gactttctctgtttattttaaacactgttacaggaaaactcgcaaaattaaattact 2262
Db 1356 taaatgatataattcttcttaaaaaaiaaaaaaaaaaaaaaaaaa 1415

Qy 2263 gaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 2311
Db 1416 aaaaaaiaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1464

RESULT 7
US-08-545-196B-10
; Sequence 10, Application US/08545196B
; Patent No. 6080577
; GENERAL INFORMATION:
; APPLICANT: MELKI, JUDITH
; APPLICANT: MUNNICH, ARNOLD
; TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
; TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-545-196B-10
```

```

Query Match      2.8%; Score 64; DB 3; Length 1582;
Best Local Similarity 65.3%; Pred. No. 3.2e-06;
Matches 94; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2168 tttaataaatatggtgggttactaaagaagtaaacgactttctctgtttattttaaca 2227
Db 1436 TTTTCTATCTCTATATGTTTAAAAAGTATATAATAAAATATTTTATTTTAAAAA 1495

Qy 2228 ctgtcagcagaactcgcaaaattaaattactgaaaaaiaaaaaaaaaa 2287
Db 1496 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1555
```


Search completed: September 6, 2002, 15:16:46
Job time: 7710 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 00:11:01 ; Search time 3580.18 Seconds
(without alignments)
8712.261 Million cell updates/sec

Title: US-09-807-470-1
Perfect score: 2311
Sequence: 1 agagagagagagagagag.....aaaaaaaaaaaaaaaaaaaaa 2311

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_gss:*
- 13: em_gss_hum:*
- 14: em_gss_inv:*
- 15: em_gss_pln:*
- 16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	717.6	31.1	825	10 BI731069	BI731069 603351703
2	699.6	30.3	773	10 BI730813	BI730813 603351811
3	628	27.2	688	9 BB648723	BB648723 BB648723
4	605.2	26.2	651	9 BB641088	BB641088 BB641088
5	594.4	25.7	645	9 BB521515	BB521515 BB521515
6	587	25.4	628	9 BB662449	BB662449 BB662449
7	577.6	25.0	610	9 BE109193	BE109193 UT-R-BS1-
8	534.4	23.1	600	10 BI985062	BI985062 3131-07 M
9	495	21.4	1356	11 BC020283	BC020283 Homo sapi
10	484.2	21.0	1064	9 AL523342	AL523342 AL523342
11	426.6	18.5	1142	10 BM466240	BM466240 AGENCOURT
12	418.2	18.1	622	12 A2958987	A2958987 2M0226C06
13	407.2	17.6	1135	10 BM466239	BM466239 AGENCOURT
14	392.4	17.0	814	9 AU132871	AU132871 AU132871
15	376.2	16.3	822	9 AU130188	AU130188 AU130188
16	373.2	16.1	790	10 BB612552	BB612552 601452065
17	370.6	16.0	944	10 BI655670	BI655670 603283957

18	369.2	16.0	905	9 AU131342	AU131342 AU131342
19	362.6	15.7	836	10 BF969571	BF969571 602271922
20	361.4	15.6	879	10 BI870787	BI870787 602394607
21	361.2	15.6	779	10 BM045930	BM045930 603624955
22	356	15.4	808	10 BI830439	BI830439 603073495
23	354.6	15.3	910	10 BM459490	BM459490 AGENCOURT
24	353.4	15.3	479	9 AI848290	AI848290 UI-M-AH1-
25	352.2	15.2	727	10 BG547017	BG547017 602573813
26	344.8	14.8	695	10 BI733351	BI733351 603353253
27	342.8	14.9	760	10 BG714759	BG714759 602673845
28	340.8	14.7	879	9 AU128453	AU128453 AU128453
29	339.4	14.7	1067	10 BF538214	BF538214 602053686
30	338.4	14.6	636	10 BG077352	BG077352 H3014E03-
31	337.6	14.6	969	10 BG975574	BG975574 602845246
32	335.6	14.5	1015	10 BG389557	BG389557 602414256
33	333	14.4	878	10 BE618241	BE618241 601462496
34	330.2	14.3	651	10 BG714581	BG714581 602671468
35	330.2	14.3	944	10 BG916945	BG916945 602816230
36	325.8	14.1	769	10 BG542329	BG542329 602571885
37	322.6	14.0	859	10 BG323904	BG323904 602422933
38	321.6	13.9	614	9 AL602739	AL602739 DKEZp686F
39	321.4	13.9	708	10 BE914578	BE914578 601665679
40	320.2	13.9	660	10 BF984410	BF984410 602307936
41	320.2	13.9	736	10 BF141338	BF141338 601789863
42	316.2	13.7	608	10 BG614318	BG614318 602642420
43	316.2	13.7	693	10 BM486565	BM486565 pgmlc.pk0
44	313.4	13.6	835	10 BE887450	BE887450 601509592
45	309.8	13.4	569	9 AI201635	AI201635 qb81d09.x

ALIGNMENTS

RESULT 1

BI731069
LOCUS
DEFINITION
mRNA sequence.
BI731069 825 bp mRNA linear EST 20-SEP-2001
603351703F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359105 5',

ACCESSION
BI731069
VERSION
BI731069.1
KEYWORDS
house mouse.
SOURCE
Mus musculus

ORGANISM
Mus musculus

REFERENCE
1 (bases 1 to 825)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov

CDNA Library Preparation: The Cepko Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM11913 row: f column: 02
High quality sequence stop: 822.

Location/Qualifiers
1. .825
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5359105"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: PCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

FEATURES
source

```
BASE COUNT      173 a      222 c      231 g      199 t
ORIGIN

Query Match      31.1%; Score 717.6; DB 10; Length 825;
Best Local Similarity 93.6%; Pred. No. 4.4e-82;
Matches 771; Conservative 0; Mismatches 49; Indels 4; Gaps 2;

Qy 380 acggcgcggtgagggagccgagcctggagcgcgcagtcagtcagcagtggtgagcgccg 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACGGCGCGGTGATGGGAGACCCCGCCCTGGAGCGGCTCAGTGATCAGCGTGGAGCGCG 60

Qy 440 ttagtcacgggtgagtcacgcgcgcgtgggctccgaagaagacccaggaaccccccgg 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TGACTACTGTTGAGTACCGCGCATGGGCTCCGCAAGAAGAACGCCAGGAACCCCGCG 120

Qy 500 tgcctagcagcaaatctatgttgcaagaccacgcgcgcgcgcgcgcgcgcgcgcgcgc 559
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 121 TGTGTAGCCACGAATTCATGTGTCAGAACACCGCGGATATGCTCTCCGTGGCATGT 180

Qy 560 tcttcgtctgggacttatgttcaggggcacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TCTTCTGCTGGGCTTATGTTGTCAGAGCACCTCCGAGATGCGATCGGTTCTTCACTC 240

Qy 620 tgcagtcagtgagtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGCAGCATGAGTCTGTTGTCACGCGGAAGGCTGCCCTCCGGTCCAGGACCTGTACC 300

Qy 680 attatggggtcaaaatctatgttgccacagtggttctctacatgctgtggccatcatc 739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 ATTATGGGGTCAAAAGATCGGCCACAGTTCCTTACATGCTGTGTGCCATCATCTCC 360

Qy 740 acgcccattcagcagtcagtcgttagataagctcagcgagcgcgcgcgcgcgcgcgcgc 799
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ACGCCACCATTCAGGAGTACGTGTAGTAAAGTACCTCAGCAGGAGACTGCAGCTCACAAG 420

Qy 800 gcaacaaacaaatctaatgagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 859
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCAACAAACAAACTGAATGAGCGCGGCGAGCTGAGTGTCTTATATAGTGTCCGGA 480

Qy 860 tctggggtgatcatctctgctctgagaaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 919
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TCTGGGGCATGATCAITCTGGCCCTTGAGAACCTGCCTGTGACACCTACACTCTTGTGGA 540

Qy 920 agtctcagcccaacacatgatgacatttcagatgaaattttctacatctcacagtgg 979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AGTCTCAGCCCCACAAACATGATGACATTCAGATGAATTTTCTTACATTTACAGATTGG 600

Qy 980 ctactggttccatagtttcccgagctctacttcagaagaagtcaggaacaaagatacc 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CTTACTGTGTTTCATAGCTTTCCGGAGCTTTACTTCCAGAAAGTCAGGAAGATATCC 660

Qy 1040 cgggtcaactcatctacattgctccacaccttccacattggagggccctatcttgt 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CGGTCAGCTCATCTACATGCGCTCCACCTCTCCACATTTGGAGGGCCCTATCTTTGT 720

Qy 1100 ac-ttgaaccacctgggctgctctctgatgctgca---ctatgctgtcgagctctc 1155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 ACTTTGAACCACTTGGGCTGCTGCTGCTGATGCTGCAACTACGCCGTTAGAGGCTCCTC 780

Qy 1156 tccagcgtgtgcagcgtcttaacttttgggagatgagcgggtacca 1199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 TCCAGCGTGTGACGCTTGCTTTATTTTGGTGTGATGAGCGGTACCA 824
```

```
RESULT      2
BI730813
LOCUS      BI730813
DEFINITION 603351811F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5359106 5',
            mRNA sequence.
ACCESSION  BI730813
VERSION    BI730813.1 GI:15707826
KEYWORDS   EST.
```

```
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 773)
AUTHORS     NIH-MGC http://mgc.mci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: The Cepko Laboratory
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            Cloning Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1913 row: f column: 03
            High quality sequence stop: 773.

FEATURES
    source
        1..773
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="IMAGE:5359106"
            /clone_lib="NIH_MGC_94"
            /tissue_type="retina"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
            Average insert size 3.3 Kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT  164 a      210 c      217 g      182 t
ORIGIN

Query Match      30.3%; Score 699.6; DB 10; Length 773;
Best Local Similarity 95.4%; Pred. No. 8.5e-80;
Matches 731; Conservative 0; Mismatches 34; Indels 1; Gaps 1;

Qy 380 acggcgcggtgagggagccgagcctgggagcgcgcagtcagtcagcagtggtgagcgccg 439
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACGGCGCGGTGATGGGAGCCCGCGGCTGGGAGCGGCTCAGTGATCAGCGTGGAGCGCG 60

Qy 440 ttagtcacgcgtgagtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 499
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 TGACTACTGTTGAGTACCGCGCATGGGCTCCGCAAGAAGAACGCCAGGAACCCCGCG 120

Qy 500 tgcctagcagcaaatctatgttgccacagtggttctctacatgctgtgggcatgt 559
    || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 121 TGTGTAGCCACGAATTCATGTGTCAGAACACCGCGGATATGCTCTCCGTGGCATGT 180

Qy 560 tcttcgtctgggacttatgttcaggggcacgcgcgcgcgcgcgcgcgcgcgcgcgcgc 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TCTTCTGCTGGGCTTATGTTGTCAGAGCACCTCCGAGATGCGATCGGTTCTTCACTC 240

Qy 620 tgcagtcagtgagtcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 679
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 TGCAGCATGAGTCTGTTGTCACGCGGAAGGCTGCCCTCCGGTCCAGGACCTGTACC 300

Qy 680 attatggggtcaaaatctatgttgccacagtggttctctacatgctgtggccatcatc 739
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 ATTATGGGGTCAAAAGATCGGCCACAGTTCCTTACATGCTGTGTGCCATCATCTCC 360

Qy 740 acgcccattcagcagtcagtcgttagataagctcagcgagcgcgcgcgcgcgcgcgcgc 799
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ACGCCACCATTCAGGAGTACGTGTAGTAAAGTACCTCAGCAGGAGACTGCAGCTCACAAG 420

Qy 800 gcaacaaacaaatctaatgagccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 859
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 GCAACAAACAAACTGAATGAGCGCGGCGAGCTGAGTGTCTTATATAGTGTCCGGA 480

Qy 860 tctggggtgatcatctctgctctgagaaactcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 919
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 481 TCTGGGGCATGATCAITCTGGCCCTTGAGAACCTGCCTGTGACACCTACACTCTTGTGGA 540

Qy 920 agtctcagcccaacacatgatgacatttcagatgaaattttctacatctcacagtgg 979
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 541 AGTCTCAGCCCCACAAACATGATGACATTCAGATGAATTTTCTTACATTTACAGATTGG 600

Qy 980 ctactggttccatagtttcccgagctctacttcagaagaagtcaggaacaaagatacc 1039
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 601 CTTACTGTGTTTCATAGCTTTCCGGAGCTTTACTTCCAGAAAGTCAGGAAGATATCC 660

Qy 1040 cgggtcaactcatctacattgctccacaccttccacattggagggccctatcttgt 1099
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 661 CGGTCAGCTCATCTACATGCGCTCCACCTCTCCACATTTGGAGGGCCCTATCTTTGT 720

Qy 1100 ac-ttgaaccacctgggctgctctctgatgctgca---ctatgctgtcgagctctc 1155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 721 ACTTTGAACCACTTGGGCTGCTGCTGCTGATGCTGCAACTACGCCGTTAGAGGCTCCTC 780

Qy 1156 tccagcgtgtgcagcgtcttaacttttgggagatgagcgggtacca 1199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 781 TCCAGCGTGTGACGCTTGCTTTATTTTGGTGTGATGAGCGGTACCA 824
```


SOURCE
ORGANISM Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
AUTHORS 1 (bases 1 to 610)
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
COMMENT Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704477
JOURNAL Contact: Soares, MB
MEDLINE Program for Rat Gene Discovery and Mapping
COMMENT University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dr track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dr track served to verify it as a clone from the
 normalized embryo at 13 dpc library cDNA library Preparation: M.B.
 Soares Lab Clone Distribution: clones will be available through
 Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
source Location/Qualifiers
 1. .610
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BS1-azd-a-04-0-UI"
 /clone_lib="UI-R-BS1"
 /dev_stage="embryonic 13 dpc"
 /lab_host="DHI0B (Life Technologies)"
 /note="vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-BS1
 library is derived from 13 dpc whole embryo tissue. For a
 detailed description of the library from which this clone
 was derived, please visit our web site at
 ratest.eng.uiowa.edu.
 TAG_LIB=UI-R-BS1
 TAG_TISSUE=embryo at 13 dpc
 TAG_SEQ=ATCC"
 206 a 118 c 95 g 190 t 1 others
BASE COUNT
ORIGIN

Query Match 25.0%; Score 577.6; DB 9; Length 610;
 Best Local Similarity 98.3%; Pred. No. 2.6e-64;
 Matches 594; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 1528 ccaagaagaagaagctccttagcagtgcaagcgaattgattcttaccctcaagg 1587
 |||||
 Db 610 CCAAGAAGAAGAAGAGAGCTCCTTAGCAGTTGCAAGCAATTGATTCTTACCTCCAGG 551
 |||||

Qy 1588 gaatccacttcttctatgtggtgtctctgtctagagatttctgtcttcagaacggg 1647
 |||||
 Db 550 GAATCCACTTCTTCTTATGTGTGTCTCTGTGCTAGAGATTTTCTGTCTTCAGAACGGG 491
 |||||

Qy 1648 tcgtgcttttgaattgtcattgtattgtcattgtttttaaagtttttgagaagt 1707
 |||||
 Db 490 TCGTGCCTTTTGAATATGCTTAATGTATGTCTTAATGTGTTTAAAGTTTTCAGACAGT 431
 |||||

Qy 1708 atgagtggggatgggggttaagactaaaccactcagcctctaataacagtcagaatagt 1767
 |||||
 Db 430 ATGAGTGGGGAGGGGGTAAAGACTAAACCACCTCAGCCTTAAACTGTTCAGAAAGT 371
 |||||

Qy 1768 taacggaccacacatcttatttagtttagttttctacccaacgattttcacaacgtttt 1827
 |||||
 Db 370 TAACGGACCAACATCTTATTAGTTAGTTCTTACCTCAACGATTTTCCAAACGTTTGT 311
 |||||

Qy 1828 gctgatgactgcagaaatgtgtacataaataatagtttctcctcttccaatgtctttatc 1887
 |||||
 Db 310 GGTGATGACTGCAGAAATGTGTACATAAATAATAGTTCCTGCTTCCAAATGTTCTTTATC 251
 |||||

Qy 1888 gaattaacaagtctgctagcaaaagtgtttgttttttctcaatgttctcctgcagataaag 1947
 |||||
 Db 250 GAATTAACAAGTCTGCTAGCAAAAGTGTGTTGTTTCTCAATGTCTCTCTGCAGATAAAG 191
 |||||

Qy 1948 tggaaaatctgataaaggttaaaactcaaatcagttattatgttaaccgttgggatttttta 2007
 |||||
 Db 190 TGGAAAATCTGATAAAGGTAAACCTCAATCAGTATATATGTAAACCGTTGGGATTTTTTA 131
 |||||

Qy 2008 aagtttttaaatattacaaatgaaagcattgttcaaacac-caaaaaatgtgtttaa 2066
 |||||
 Db 130 AAGTGTTTTAAATTTACAAATGGAAGCATTTGTCAAAACCCACNCAAAATATGTGTTAAT 71
 |||||

Qy 2067 ttatgagtagtaattgttagtgcttacccccccattaaagcatcaaaatagaatagat 2126
 |||||
 Db 70 TTTATGAGTAGTAATTGTTAGTGTCTTACGCCCATTAAGCATCAAAATTTGAAAAAAA 11
 |||||

Qy 2127 gaca 2130
 |||
 Db 10 AAAA 7
 |||

RESULT 8
BI985062
LOCUS BI985062 600 bp mRNA linear EST 20-DEC-2001
DEFINITION 3131-07 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
 mRNA sequence.
ACCESSION BI985062
VERSION BI985062.1 GI:179555990
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 600)
AUTHORS Mu X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
TITLE Gene expression in the developing mouse retina by EST sequencing
 and microarray analysis
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)
COMMENT Contact: Klein WH
 Department of Biochemistry and Molecular Biology
 University of Texas M.D. Anderson Cancer Center
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
 Tel: 713 792 3646
 Fax: 713 790 0329.

FEATURES
source Location/Qualifiers
 1. 600
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
 /tissue_type="neural retina"
 /dev_stage="embryonic day 14.5 post-fertilization"
 131 a 159 c 168 g 140 t 2 others
BASE COUNT
ORIGIN

Query Match 23.1%; Score 534.4; DB 10; Length 600;
 Best Local Similarity 95.3%; Pred. No. 7.8e-59;
 Matches 572; Conservative 0; Mismatches 26; Indels 2; Gaps 2;

Qy 409 gagcgcgcagtgatcagcgggtggcgccggtgagtcacccggtgagtcacccgcgcagatgggg 468
 |||||
 Db 1 GAGCGCTCAGTGATCAGCGGTGAGCGCGGTGACTACTGTTGTTGAGTACCGCGCATGGGG 60
 |||||

Qy 469 ctcccagaagaacgcccaggaaccccccggtgtgtgtagcac-caattcatggtgcagaa 527
 |||||
 Db 61 CTCCGACGAGAGAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
 |||||

```
QY 528 ccacgc-gaatagtctctcgtgtaggcatgtcttctgctgggacttatgtcagg 586
Db 121 CCACGCGGATATGGTCTCTCGGTGGGATGTTCTTCTGCTGGGCTTATGTTTGAGG 180
QY 587 gcacgcccagagatgtcgatcgtgttctcaccctcagcatgagtcgttgtccccagcg 646
Db 181 GCACCTCCGAGATGTCGATCGGTTCTCTACCTCTGCAGCATGGAGTCGTTGTGCCAGCG 240
QY 647 aggggtacccctcgggggtccagagaccctttaccattatgggggtcaaatgtctggcacag 706
Db 241 AAGGGTGCCTCCGGGTCCAGGACCTGTACCATTTATGGGGTCAAGATCTCGGCCACAG 300
QY 707 tgtttctacatgctggtggccatcatctcaccacccattcaggagtagctgtag 766
Db 301 TGTCTCTACATGCTGGTGGCCATCATATCCAGCCACCATTCAGGAGTAGCTGCTAG 360
QY 767 ataagctcagcggagactgcagctcaccacaaagcgaacacaaattgaatgagccg 826
Db 361 ATAAGCTCAGCAGAGACTGACGCTCACCAGGCAACAAACAACTGAATGAGGCGC 420
QY 827 ggcagctgagtggtctctacatagtgctgtgtatctcgggtatgatcttctgacctcg 886
Db 421 GGCAGCTGAGTGTCTTATATAGTGTCCGGGATCTGGGCGATGATCATTTCTGGCCCTG 480
QY 887 agaacgtcgtcagaccccaactctattgtggaagtctcagcccaacacatgatgacat 946
Db 481 AGAAGCTGCTGTGACACCTTACACTCTGTGGAAGTCTCAGCCGCCAACATGATGACAT 540
QY 947 ttcagatgaattttctacatctcagctcagttggtctactggtttcattagttccccgagc 1006
Db 541 TTCAGATGAATTTTCTACATTTTACAGTTGGCTTACTGGTTTCTAGCTTCCGGAGC 600

RESULT 9
BC020283 1356 bp mRNA linear HTC 20-DEC-2001
LOCUS Homo sapiens, clone IMAGE:4519858, mRNA.
DEFINITION BC020283
ACCESSION BC020283.1 GI:17946782
VERSION HTC.
KEYWORDS Homo.
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1356)
Strausberg, R.
Direct Submission
Submitted (20-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Rulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R.,
Lawrence, S., Richards, S., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 29 Row: h Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
```

This clone has the following problem: no 5' EST match.

```
FEATURES
    source          Location/Qualifiers
     1..1356
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:4519858"
        /tissue_type="Prostate, adenocarcinoma."
        /clone_lib="NIH_MGC_91"
        /lab_host="DH10B"
        /note="Vector: pCMV-SPORT6"
BASE COUNT      401 a   279 c   268 g   408 t
ORIGIN

Query Match      21.4%; Score 495; DB 11; Length 1356;
Best Local Similarity 65.0%; Pred. No. 6.e-54;
Matches 825; Conservative 0; Mismatches 410; Indels 34; Gaps 5;

QY 436 ccggtgagtagcggtagtaccgcgccatgggggtccgcaagaagaacccaggaacccc 495
Db 49 CCTGCAGCGCGGGCGCGGCCCCACCATGGCGATTTCGAAGAAAAAGCACCAGAGCCCC 108
QY 496 ccgggtgcagcagaattcatggtgcagaaccacgcgcatatggtctcctcgtcgtggc 555
Db 109 CCAGTGTGTGAGCCAGAAATTCGTCTCGAAGATCACGGGACATCGTCTCTGTGTGGCG 168
QY 556 atgttctcgtcgtgggacttatgttcgagggccagccgagagatgctcgtcgttcctc 615
Db 169 ATGGTCTCTCTGCTGGGGCTCATGTTTGAGATAAAGGCAAAAGCTTCTATCATTTTGT 228
QY 616 accctgcagcatggagtcgttgtcccgag ---cggaaagggtaccctcgggggtccaggacc 672
Db 229 ACTCTTCAGTACAATGTCCACCTCCACGCAACAGAAAGCAAGCTTACTGAATCAGTGTCC 288
QY 673 ctttaaccattatgggtcaaatgctgcacagtggtcttctctacatcgtgtggccatc 732
Db 289 CTTTATTACTATGCGATCAAAAGATTTGGCTACTGTTTCTTCTTCTACATGCTAGTGGCGATA 348
QY 733 atcattcacgcaccattcaggagtagctgtctagataagctcagccgagagactgcagctc 792
Db 349 ATTATTTCATGCCGTAATTTCAAGAGTATATGTTGGATAAAATTAACAGCGGAATGCATTC 408
QY 793 accaaaggcaaaaaaacaattgaatgagccggcgagctgagtggtttctacatagtg 852
Db 409 TCCAAAAACAAACACAGCAAGCTGAATGAATGTGCTGAGCTTAGTGCGTCTTACCTTTT 468
QY 853 tctggtatctgggtgatgatcattctgcctctgagaagctgcctgcagacccccactcta 912
Db 469 GTCTGTGTTTGGGGCACATTTCATTCATCTCTGAAGAACTACATCTCAGACCCCAACTATC 528
QY 913 ttgtggaagtctcagcccccacacatgatgacatttcagatgaataattttctacatctca 972
Db 529 TTATGGAGGGCTTATCCCCATAAACCTGATGACATTTCAAAATGAAGTTTTTCTACATATCA 588
QY 973 cagttggcttactggtttcattagtttccggagctctacttccagaagaagtcaggaaacaa 1032
Db 589 CAGCTGGCTTACTGGCTCTATGCTTTTCTGTAACCTTACTTCCAGAAAACCAAAAAGAA 648
QY 1033 gatattccgggtgtaactcatctacattggcctccaccctcttccacattggaggggacctat 1092
Db 649 GATATCTCTCGTCAAGCTTGTCTACATTTGGTCTTTTACCTCTTCCACATTCCTGGAGCTTAC 708
QY 1093 cttctgtactgaacacacactgggacctgctgcttctcgtatgctgacactatgctcagacct 1152
Db 709 CTTTGAAGCTTTGAATCATCTAGGACTTGTCTCTGGTGTCTACATTTTGTGTGAATTT 768
QY 1153 ctctccagcgtgtagcctgcttactttgggtggagcggtagccagaaagggtgtctc 1212
Db 769 CTTTTCACATTTCCGCGCTGTTTATTTTATTCATGAAGAAAGTATCAGAAAGGATTTCT 828
QY 1213 ttgtggcctatcgtgttttatatccggggagactcgtgacactgattgctcagtggttaca 1272
Db 829 CTGTGGCGAGTCTCTTTTGTGTTGGGAGAGACTTCTGACTTTAATTAATCTTCTCAGTACTGACT 888
```

```

QY 1273 gtaggcttcacttgcccgagac---aaatcggaatgaaatgctctctgtaatgct 1329
Db 889 GTTGGTTTGGCCCTTCAGAGCAGGAGAAATCAGAACTGGATTTTCAGTACTGAAACTTC 948
QY 1330 aatgtgttgccagcagtaaatcgctgttctgctcgcagttgcagttaccaggtgtacata 1389
Db 949 AATGTGTTAGCTGTTAGAAATCGCTGTTCTGGCATCCATTTGCGTTTACTCAGGCATTTATG 1008
QY 1390 aatggacottgacaccgtctgcttcagagatggttgaagaatgcgaatcttcattgctc 1449
Db 1009 ATGTGGAAGTTCAATTAATTTTCAGCTTCGAAGGTGAGGGAACATTTCTGTTTTCAGGCA 1068
QY 1450 tggggaggaag-----agacggtccaggtcgagaaagccacagaa 1491
Db 1069 CCAGCTGTGAGAGAAACCAACAGTAACATAAGGCAGATCTTCTAATAAAGGAACAGAA 1128
QY 1492 aatggagtgga-----gaatccaaatagaaatgattctccacaaagaagaagag 1542
Db 1129 AATGGTGTGAATGGAACATTAACCTCAAAATGTAGCAGACTCTCCCGGAATAAAAAAGAG 1188
QY 1543 aaagc-tccttagcaattgcaagcgaattgattcttacctccaagggaatccactcttc 1601
Db 1189 AAATCTTCATAATGAATTAATAACTAATTAATGAATTAATGTCCTCCCAAGAAATCTGCTTTCTA 1248
QY 1602 ttatgtggtctctgctgtagagatttctgttcttcagacgggtcggtgctttttgaa 1661
Db 1249 CTATATCTTTCAGCATTAGAGATTTTCTGTTCTTGAAATATACAGTCTGTGCTTTTGAT 1308
QY 1662 tattgctaa 1670
Db 1309 TTTTAAAAA 1317

RESULT 10
AL523342
LOCUS
DEFINITION AL523342 LFI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC001YH18 5
prime, mRNA sequence.
ACCESSION AL523342
VERSION AL523342.1 GI:12786835
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1064)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. 1064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DC001YH18"
/clone_lib="LFI_NFL003_NBC3"
/sex="male"
/tissue.type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

```

```

BASE COUNT 270 a 236 c 220 g 336 t 2 others
ORIGIN
Query Match 21.0%; Score 484.2; DB 9; Length 1064;
Best Local Similarity 68.7%; Pred. No. 1.6e-52;
Matches 695; Conservative 1; Mismatches 309; Indels 6; Gaps 2;
QY 426 gcggtgcccgcggtgagtaccggtgagtcacggtgggtccgcagaagaacgc 485
Db 1 GGGCGSASACCTTCAGCGCGGGCGGCCACCATGGCGATTCGCAAGAAAGCAC 60
QY 486 cagaaaccccccggtgctgagccacgaattcatggtgcagaacacgcggtatggtctc 545
Db 61 CAAGAGCCCCCAGTGTGAGCCACGAATTCGTCTGCAGAAATCACGGCGACATCGTCTC 120
QY 546 ctgcgtgggcatgtcttcgtggtggaacttatgttcgagggcgcgagatgctgat 605
Db 121 CTGTGTGGCATGTGCTTCTCTGCTGGGCTCATGTTTGAGATAACGGCAAAAGCTTCAT 180
QY 606 cgtgttctcacccctgcagcatggagtcgtgtgtccag---cgaaagggtaccctcggg 662
Db 181 CATTTTGTGTTACTCTTCAGTACAATGTCAACCTCCCAACAACAGAAAGACTACTGA 240
QY 663 gtccaggacccctttaccattatgggttcaagatctggccacagtggtcttctcatgct 722
Db 241 ATCAGTGTCCCTTTATTACTATGTCATCAAGATTTGGCTACTGTTTCTTCTACATGCT 300
QY 723 ggtggccatcatcattcacgccaccattcaggtgacgtacgtgtagatagtcagccgag 782
Db 301 AGTGGCGATAAATTATTCATGCGGTAATTCAAGAGTATATGTTGGATAAAATTAACAGCG 360
QY 783 actgcagctcaaaaggcaaaacaaataagacggccgagcggcgagtggtgt 842
Db 361 AATGCACCTCTCCAAAACAAACACAGCAAGTTTAATGAATCTGTCAGCTAGTGCCTT 420
QY 843 ctacatagtgctgtgattctgggttatgattcatcttctgacctctgagaactgcctgcaga 902
Db 421 CTACCTTTTTCCTGTGTTGGGCGACATTCATTTCTCATCTCTGAAAACATACATCTCAGA 480
QY 903 cccactctattgtggaagtcctcagcccccacacatgatgaatttcagatgaattttt 962
Db 481 CCCAACTATCTTATGGAGGGCTTATCCCATCAACCTGATGACATTTCAAAATGAAGTTTTT 540
QY 963 ctacatctcacagttgcttactggtttcatagtttccggagctctacttcccaaaagt 1022
Db 541 CTACATATCATCAGCTGGCTTTACTGGCTTCTGCTTTTCTGAACTCTACTCTCCAGAAAC 600
QY 1023 cagaaacaagatatcccggttcaactcatctacattggcctccacctcttccacattgg 1082
Db 601 CAAAAAAGAGATATTCCTCGTCAGCTTGCTACATTGGTCTTTTACCTTCCCATTTGC 660
QY 1083 aggggcctatcttctgttacttgaaccacctgggcctgctgtcttctgtatgctgcactatgc 1142
Db 661 TGGAGCTTACCTTTTGAACCTTGAATCATCTAGGACTTGTCTCTGCTGCTACATATTAT 720
QY 1143 tgtcagctcctctccagcgtgtgcagcctcttacttcttcttgggagatgagcggtaccagaa 1202
Db 721 TGTGTAATTTCTTTTCCACATTTCCCGCTGTTTTTATTTAGCAATGAAAGAAATATCAGAA 780
QY 1203 aggggttcttctgtgacctatcggttttatatccgggagacctcgtagcactgattgtctc 1262
Db 781 AGGATTTTCTCTGTGGGCGAGTCTTTTGTGTTTGGGAAGACTTCTGACTTTTAATTTCTTC 840
QY 1263 agtggttacagtaggggttctacttggccgggac---aaatcggaatggaatgctctctc 1319
Db 841 AGTACTGACTGTTGGTTTGGCTTGCATTGCAAGACAGAAAAATCAGAAAGCTGGATTTTCAGTAC 900
QY 1320 tggtaatgtaaatgtgttggcagctaaatcgctgttctgctcctcagattcagtatcca 1379
Db 901 TGGAAACTTCAATGTGTAGCTGTAGAACTGCTGTTTGGCATCCATTTGCGTTACTCA 960
QY 1380 ggtgtacataacatggaccttgacgacctctggtgcttcagagatggttaga 1430

```


Db 506 TTATGGAGGCGCTTATCCCAATACCTGATGACATTTCAAAATGAAGTTTTTCTACATATCA 565
QY 973 cagttgcttactgtttccatagtttccggagctctacttccagaagtccagaaacaa 1032
Db 566 CAGCTGGCTTACTGGCTTTTCATGCTTTTCTGAACTCTACTTCCAGAAAACCCAAAAAGAA 625
QY 1033 gataaccgggtcaactcatctacatgttgctccactctccacattggagggcctat 1092
Db 626 GATATTCCTCGTCAGCTTGTCTACATTTGCTTTTACTCTTCCACATTGCTGGAGCTTAC 685
QY 1093 ctctgttacttgaacacacctggccctgctgctctgtatgctgactatgctgcagctc 1152
Db 686 CTTTGAACCTGAATCATCTAGGACTTGTCTTCTGCTGCTACATTTATTTGTTGAATT 745
QY 1153 ctctccagctgtcagctcgttacttggggt-gagcgtaccagaaagg---tt 1208
Db 746 CTTTCCACATTTTCGGCTGTTTATTTTACCAATGGAAGATACAGAAGGGGATTTT 805
QY 1209 gtcttgtggcctatgctgtttatatccgggagactcgtgacactgattgctcagtggt 1268
Db 806 CTTCTGGGCCAGNCTCTTTTGTGTTTTTGGGAAACCTCTGACTTTTAATCCTTTCAGAACT 865
QY 1269 t 1269
Db 866 T 866

RESULT 14
AUI32871
LOCUS AUI32871 NT2RP4 Homo sapiens cDNA clone NT2RP4000757 5', mRNA EST 24-OCT-2000
DEFINITION
ACCESSION AUI32871
VERSION AUI32871
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,
Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and
Isogai,T.
HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y.,
Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano
'S., Masuho,Y., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute
Location/Qualifiers
1. .814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4000757"
/clone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

BASE COUNT 206 a 195 c 161 g 249 t
ORIGIN

Query Match 17.0%; Score 392.4; DB 9; Length 814;
Best Local Similarity 70.7%; Pred. No. 7.5e-41;
Matches 550; Conservative 0; Mismatches 223; Indels 5; Gaps 2;
QY 436 ccggtgagtagtaccggtgagtagtaccggtggtccgcgaagaagaagccaggaacccc 495
Db 18 CTTGTCAGGGCGGGCGGGCCACCATGGCATTTCCAAAGAAACGACCAAGAGCCCC 77
QY 496 ccggtgtagccagccagaaattcgtgagaaacacgcgcgatatggttctctcgtcggtgc 555
Db 78 CCAAGTCTGAGCCACAGAAATTCGTCGCAAGAAACACGGGACATCGTCTCTGTGTGGCG 137
QY 556 atgtcttctgctgggacttatgttcgagggcagccgagatgtcgatgtcttctc 615
Db 138 ATGGTCTTCTGCTGGGGCTCATGTTTGAGATAACGCAAAAGCTTCTATCATTTTGT 197
QY 616 accctcagcatgagtcggtgtcccaag---cggaaaggctacccctcggtgtccagacc 672
Db 198 ACTCTTCAGTACAAATGTACCCCTCCCGACCAAGAAAGTACTTGAATCAGTGTC 257
QY 673 ctttacccattatgggtccaaagatctgtgccacagtgcttctctacatgctgtgtgccatc 732
Db 258 CTTTATTACTATGGCATCAAGATTTGGCTACTGTTTTCTTCTACATGCTAGTGGCGATA 317
QY 733 atcattccagccaccattcaggagtagctgtagataaagctcagccgagagactcagctc 792
Db 318 ATTATTTCATGCCGTAATTCAAGAGTATATGTTGGATAAAATTAACAGCGCAATGCAC 377
QY 793 accaaaggccaaacaaattgaatgagccggcgagctgagctgtgtctacatagtg 852
Db 378 TCCAAAACAAAACACAGCAAGTTTAAATGAATCTGGTCAGCTTAGTGGCTTCTACCTTTT 437
QY 853 tctggtatctggggtatgatactctgtgcctcgagaagtcgctgtcagaccctactcta 912
Db 438 GCCTGTGTTTGGGGCACATTCATCTCTCTGAAAACACTACATCTCAGACCCCAACTC 497
QY 913 ttggaagtctcagcccccacacatgatgacatttcagatgaaatttttttctacatctca 972
Db 498 TTATTGGAGGCTTTATCCCATAACTGATGACATTTCAAATGAAGTTTTTCTACATATCA 557
QY 973 cagttggttactggtttctaatgtttcccgagctctacttccagaagtcagaaacaa 1032
Db 558 CAGCTGGCTTACTTGGCTTTCATGCTTTTCCGAACTCTACTTCCAGAAAACCCAAAAAGAA 617
QY 1033 gatataccgggtcaactcatctacattggcctccacctcttccacatttggagggcctat 1092
Db 618 GATATTCCTCGTCAGCTTGTCTACATTTGGTCTTTACCTCTTCCACATTTGCTGGAGCTTAC 677
QY 1093 ctctgttacttgaacacacctggcctgctcttctgtatgctgcactatgctgcagctc 1152
Db 678 CTTTGTGAACCTGAATCATCTAGGACTTGGTCTTCTGGTGTACATTTAT--TTGTGAATTT 735
QY 1153 ctctccagctgtgcagctgcttacttggggatgagcggtaccagaaagggttgt 1210
Db 736 CTTTTCACATTTCCCGCCTGGTTTATTATTAGCANTGAAAAGTATCAGAAANGGATTT 793

RESULT 15
AUI30188
LOCUS AUI30188 NT2RP3 Homo sapiens cDNA clone NT2RP3000388 5', mRNA EST 24-OCT-2000
DEFINITION
ACCESSION AUI30188
VERSION AUI30188
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 822)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.

— — — — —

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 14:51:51 ; Search time 74.4 seconds
(without alignments)
541.933 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKNARNPVLSEFMV.....NGVENPNRIDSPKKKKKAP 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1892	100.0	363	21	AA198146		Rat WAR-1 amino ac
2	1892	100.0	363	22	AA198146		Rat WAR-1 protein
3	1463	77.3	369	21	AA198147		Human WAR-1 amino
4	1463	77.3	369	22	AB12234		Novel human diagno
5	1463	77.3	369	22	AA198146		Human WAR-1 protein
6	1462	77.3	369	22	AA198146		Human polypeptide,
7	1134.5	60.0	416	21	AA143601		Human cancer assoc
8	922	48.7	304	20	AA148434		Human prostate can
9	739	39.1	384	22	AA195708		Human reproductive
10	566	29.9	368	22	AB165847		Drosophila melanog
11	566	29.9	368	22	AB166114		Drosophila melanog

12	566	29.9	1575	22	AB167362	Drosophila melanog
13	414.5	21.9	125	21	AA198146	Human secreted pro
14	148.5	7.8	411	17	AA198146	Saccharomyces cere
15	138.5	7.3	394	22	AA198146	Human LAPH-1 prote
16	138.5	7.3	394	22	AA198146	Human protein sequ
17	138.5	7.3	394	22	AA198146	Human protein sequ
18	133.5	7.1	191	21	AA198146	Human secreted pro
19	117	6.2	488	22	AA198146	Human polypeptide
20	117	6.2	536	22	AA198146	Human protein sequ
21	115.5	6.1	286	21	AA198146	Arabidopsis thalia
22	114.5	6.1	280	21	AA198146	Arabidopsis thalia
23	111.5	5.9	158	22	AA198146	Novel signal trans
24	111.5	5.9	181	22	AA198146	Novel signal trans
25	111	5.9	312	22	AA198146	Human olfactory re
26	111	5.9	312	22	AA198146	Human OR-like poly
27	110	5.8	380	20	AA198146	Human LAPH-2 prote
28	110	5.8	380	22	AA198146	Human hydrophobic
29	108	5.7	311	22	AA198146	Human apoptosis pr
30	106.5	5.6	266	21	AA198146	Human cancer assoc
31	105.5	5.6	310	21	AA198146	Arabidopsis thalia
32	101.5	5.4	150	21	AA198146	Human secreted pro
33	98	5.2	232	22	AA198146	Drosophila melanog
34	97.5	5.2	313	21	AA198146	zea mays protein f
35	97.5	5.2	397	20	AA198146	Chlamydia trachoma
36	97.5	5.2	533	22	AA198146	Human polypeptide
37	96.5	5.1	431	22	AA198146	Novel human diagn
38	96	5.1	254	21	AA198146	Arabidopsis thalia
39	93	4.9	313	21	AA198146	Human G protein-co
40	93	4.9	313	22	AA198146	Human olfactory re
41	92	4.9	313	22	AA198146	Human olfactory re
42	91	4.8	307	22	AA198146	Human olfactory re
43	91	4.8	315	22	AA198146	Human olfactory re
44	91	4.8	324	22	AA198146	Human olfactory re
45	90	4.8	309	22	AA198146	G-protein coupled

ALIGNMENTS

RESULT	1
AA198146	AA198146 standard; Protein; 363 AA.
ID	AA198146
AC	AA198146
DT	22-AUG-2000 (first entry)
XX	Rat WAR-1 amino acid sequence.
DE	Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
XX	diagnosis; cancer; sarcoma; rat.
KW	Rattus norvegicus.
XX	WO200022123-A1.
PN	20-APR-2000.
XX	13-OCT-1999; 99WO-JP05631.
PF	13-OCT-1998; 98JP-0290711.
XX	(SUMU) SUMITOMO PHARM CO LTD.
PA	Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX	Okuyama H;
PI	WPI: 2000-317980/27.
XX	N-PSDB; AAA38032.
DR	Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
XX	proliferation for use in treatment and diagnosis of cancer including
PT	sarcomas of high malignancy -

XX
PS Claim 1; Fig 2; 89pp; Japanese.
XX
CC This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridize to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
SQ Sequence 363 AA;

Query Match 100.0%; Score 1892; DB 21; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.8e-213;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRKKARNPPVLSHEPMQNHADMVSCVGMFFVLGLMFEFTAEMSIIVTLTQHGVVVP 60
Db 1 mglrkkarnppvlshemvqnhadmvcvgmffvlgmfeftaemsivfittlqhgvvvp 60
QY 61 AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQTKGQNKLINE 120
Db 61 aeglpssgtrtlyhygvkdlatvffymvliiaiihataiqeyvldksrrlqtkgqnkline 120
QY 121 AGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSQPHNMFTFQMKFFYISQLAYWFHSFP 180
Db 121 agqlsvfyivsgiwgmilaseclsdptllwksqphnmmtfqmkffiyisqlaywfhsfp 180
QY 181 ELYFQKVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYF 240
Db 181 elyfqvrkqdipgqliyiglhfihgagayllylnhlglllmlhyavellssvcsllf 240
QY 241 GDERYQKGLSIWPIVFIISGRVLTIVSVTVGLHLAGTNRNGNALSGNVNLAATAVLVS 300
Db 241 gderqkglslwpivfisgrvltivsvtvglhlagtnrngnalsgnvnlavlaavlvs 300
QY 301 SSCSIQYIITWTTLTVWLQWLEADANLHVCGRKRRSRKGTENGVENPNRIDSPPKKKE 360
Db 301 sscsiqyitwtlttvwlqrlwledanlhvcgrkrrsrkgtengvenpnridspppkke 360
QY 361 KAP 363
Db 361 kap 363

RESULT 2
AAB70696
ID AAB70696 standard; Protein; 363 AA.
AC AAB70696;
XX
DT 17-MAY-2001 (first entry)
DE
DE Rat WAR-1 protein sequence SEQ ID NO:2.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
OS
OS Rattus sp.
XX
PN WO200114582-A1.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-JP05488.
XX
PR 20-AUG-1999; 99JP-0234764.
XX

PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Okuyama H, Inamura M, Ishikawa H, Nemoto K;
XX
DR WPI: 2001-202940/20.
DR N-PSDB; AAF74781.
XX
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
XX
PS Claim 3; Page 60-62; 79pp; Japanese.
XX

CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC represents a specifically claimed rat WAR-1 protein from the present
CC invention.

SX Sequence 363 AA;

Query Match 100.0%; Score 1892; DB 22; Length 363;
Best Local Similarity 100.0%; Pred. No. 3.8e-213;
Matches 363; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRKKARNPPVLSHEPMQNHADMVSCVGMFFVLGLMFEFTAEMSIIVTLTQHGVVVP 60
Db 1 mglrkkarnppvlshemvqnhadmvcvgmffvlgmfeftaemsivfittlqhgvvvp 60
QY 61 AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQTKGQNKLINE 120
Db 61 aeglpssgtrtlyhygvkdlatvffymvliiaiihataiqeyvldksrrlqtkgqnkline 120
QY 121 AGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSQPHNMFTFQMKFFYISQLAYWFHSFP 180
Db 121 agqlsvfyivsgiwgmilaseclsdptllwksqphnmmtfqmkffiyisqlaywfhsfp 180
QY 181 ELYFQKVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYF 240
Db 181 elyfqvrkqdipgqliyiglhfihgagayllylnhlglllmlhyavellssvcsllf 240
QY 241 GDERYQKGLSIWPIVFIISGRVLTIVSVTVGLHLAGTNRNGNALSGNVNLAATAVLVS 300
Db 241 gderqkglslwpivfisgrvltivsvtvglhlagtnrngnalsgnvnlavlaavlvs 300
QY 301 SSCSIQYIITWTTLTVWLQWLEADANLHVCGRKRRSRKGTENGVENPNRIDSPPKKKE 360
Db 301 sscsiqyitwtlttvwlqrlwledanlhvcgrkrrsrkgtengvenpnridspppkke 360
QY 361 KAP 363
Db 361 kap 363

RESULT 3
AAY98147
ID AAY98147 standard; Protein; 369 AA.
XX
AC AAY98147;
XX
DT 22-AUG-2000 (first entry)
XX

DE Human WAR-1 amino acid sequence.
XX
KW Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; human.
OS Homo sapiens.
XX
PN WO200022123-A1.
XX
XX 20-APR-2000.
XX
XX 13-OCT-1999; 99WO-JP05631.
XX
XX 13-OCT-1998; 98JP-0290711.
XX
XX (SUMI) SUMITOMO PHARM CO LTD.
XX
XX Tachdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
XX
DR WPI; 2000-317980/27.
DR N-PSDB; AAA38013.
XX
XX Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
PT proliferation for use in treatment and diagnosis of cancer including
PT sarcomas of high malignancy
PT
XX
XX Claim 1; Fig 2; 89pp; Japanese.
XX
XX This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridise to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
SQ Sequence 369 AA;

Query Match 77.3%; Score 1463; DB 21; Length 369;
Best Local Similarity 75.8%; Pred. No. 9.6e-163;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKKNARNPPVLSHEFMVONHADMSCVGMFFVLGLMFEFTAEMSVIFLTLOHGVVVP 60
DB 1 mglrkkstknppvlsqefilqnhadivscvgmffilgvlfgtaeasivfltiqnsavp 60
QY 61 -AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLOLTGKQKNLN 119
DB 61 aaeqatgsksllyygvkdlatvffymvaihathiqeyvldkinkrmftkakqkfn 120
QY 120 EAQQLSVFYTVSGIWMIIASCNCLSDPILLWKSQPHNMFTQMKFFYISQLAYWPHSF 179
DB 121 esqgfsyvfisfclwtgffiliscnclsdptliwkarphsmntfmgkffysqlaywfhaf 180
QY 180 PELYFQVKRKQDIPGQLIYGLHFLHFGGAYLLYLNHLGLMLHVAVELLSVCCLLY 239
DB 181 pelyfaktkqdlprqlvylglnhflhtgayillynhlgilvlyhfvellshmcglfy 240
QY 240 FGDERYQKGLSLWPIYFISGRVLTIVSVVTVGLHLAGT-NRNGNALSGNVNVLAAKIAV 298
DB 241 fsdekyqkglslwaivfigrvltivsvitvqfhlagsqnrpdaltgnvnlakiaiv 300
QY 299 LSSCSIQYITWTLTVWQWLEDANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
DB 301 lsscsitqayvtwnltlwlqrvvedsnlqacmkkkrsskrktengvgvetsnrvc 360
QY 355 PPKKKEKA 362
DB 361 ppkkrkes 368

RESULT 4
ABG12234
ID ABG12234 standard; Protein; 369 AA.
XX
AC ABG12234;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12225.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS76421.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 42593; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 369 AA;

Query Match 77.3%; Score 1463; DB 22; Length 369;
Best Local Similarity 75.8%; Pred. No. 9.6e-163;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKKNARNPPVLSHEFMVONHADMSCVGMFFVLGLMFEFTAEMSVIFLTLOHGVVVP 60
DB 1 mglrkkstknppvlsqefilqnhadivscvgmffilgvlfgtaeasivfltiqnsavp 60
QY 61 -AEGLPSSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLOLTGKQKNLN 119
DB 119

Db 61 aaeqatgsksllyygvkdlatvffymlvaiiiahatiqeyvldkinkrmqftkakqkfn 120
 QY 120 EAGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSOPHNMMTFQMKFFYISOLAYWFHGF 179
 Db 121 esgqfsvfyfsciwgtfllisencslsdptllwkarphsmmtfcmkffysqlaywfhaf 180
 QY 180 PELYFQVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLMLHYAVELLSSVCSLLY 239
 Db 181 pelyfqtktkqdiqrlyvylghlhitgayllylnhlglllvlyfvelshmcglfy 240
 QY 240 FGDERYQKGLSWPIVIFISGRVLTVLVSVTVGLHLAGT-NRNGNALSGNVNVLAAKIAV 298
 Db 241 fsdekyqkgsiwaivfllgrivtlivsvltvgfhlagsgnnpdaltgnvnvlaakiav 300
 QY 299 LSSSCSIQVYITWTLTVMLQWLEDANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
 Db 301 lssscicqayvtnliltwlrwvedsnlqascmkkrssrskrtengvgvetsnrvc 360
 QY 355 PPKKKEKA 362
 Db 361 ppkrkeks 368
 RESULT 5
 ID AAB70695 standard; Protein; 369 AA.
 XX
 AC AAB70695;
 XX
 DT 17-MAY-2001 (first entry)
 XX
 DE Human WAR-1 protein sequence SEQ ID NO:1.
 KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
 KW endoplasmic reticulum membrane transportation; secretory protein;
 KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
 KW antirheumatic; nervous system disorder; immune disorder; allergy;
 KW rheumatism; skeletal disorder.
 OS Homo sapiens.
 XX
 PN WO200114582-A1.
 XX
 PD 01-MAR-2001.
 XX
 PF 17-AUG-2000; 2000WO-JP05488.
 XX
 PR 20-AUG-1999; 99JP-0234764.
 XX
 PA (SUMI) SUMITOMO PHARM CO LTD.
 XX
 PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
 XX
 DR WPI; 2001-202940/20.
 DR N-PSDB; AAF74780, AAF74782.
 XX
 PT Transformation of a cell with separate vectors expressing the sense and
 PT antisense strands of WAR-1 DNA for screening secretory and membrane
 PT proteins expressed by the cell -
 XX
 PS Claim 2; Page 58-60; 79pp; Japanese.
 XX
 CC The present invention describes a screening method for secretory and
 CC membrane proteins consisting of transformation of a cell with separate
 CC expression vectors for the sense and antisense RNA of DNA encoding an
 CC endoplasmic reticulum membrane protein participating in endoplasmic
 CC reticulum transport of proteins. Also described are: (1) secretory and
 CC cell membrane proteins identified by the screening method; (2) drug
 CC compositions containing these proteins; (3) host cells transformed by
 CC the separate expression vectors of the method; and (4) the preparation
 CC of secretory and cell membrane proteins by culture of the transformants.
 CC The method can be used for the identification and preparation of
 CC proteins for use in the treatment and prevention of diseases such as

CC cancer, disorders of the nervous system, immune disorders (including
 CC allergies and rheumatism) and skeletal disorders. The present sequence
 CC represents a specifically claimed human WAR-1 protein from the present
 CC invention.
 XX
 SQ Sequence 369 AA;
 Query Match 77.3%; Score 1463; DB 22; Length 369;
 Best Local Similarity 75.8%; Pred. No. 9,6e-163;
 Matches 279; Conservative 43; Mismatches 103; Indels 6; Gaps 5;
 QY 1 MGLRKKNARNPPVLSHEFMVONHADMVSCVGMFVLCIMFEGTAEMSIVFLTLQHGYYVP 60
 Db 1 mglrkstknppvlsqefllqnhadivscvgmffllglvfegetaaasivfltiqhsavp 60
 QY 61 -ABGLPSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRLQLTGKQKNLN 119
 Db 61 aaeqatgsksllyygvkdlatvffymlvaiiiahatiqeyvldkinkrmqftkakqkfn 120
 QY 120 EAGQLSVFYIVSGIWMGMIILASENCLSDPTLLWKSOPHNMMTFQMKFFYISOLAYWFHGF 179
 Db 121 esgqfsvfyfsciwgtfllisencslsdptllwkarphsmmtfcmkffysqlaywfhaf 180
 QY 180 PELYFQVRKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLMLHYAVELLSSVCSLLY 239
 Db 181 pelyfqtktkqdiqrlyvylghlhitgayllylnhlglllvlyfvelshmcglfy 240
 QY 240 FGDERYQKGLSWPIVIFISGRVLTVLVSVTVGLHLAGT-NRNGNALSGNVNVLAAKIAV 298
 Db 241 fsdekyqkgsiwaivfllgrivtlivsvltvgfhlagsgnnpdaltgnvnvlaakiav 300
 QY 299 LSSSCSIQVYITWTLTVMLQWLEDANLHV-CGRKRSR-SRKGTEN--GVENPNRIDS 354
 Db 301 lssscicqayvtnliltwlrwvedsnlqascmkkrssrskrtengvgvetsnrvc 360
 QY 355 PPKKKEKA 362
 Db 361 ppkrkeks 368
 RESULT 6
 ID AAM93265 standard; Protein; 369 AA.
 XX
 AC AAM93265;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 2725.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 OS Homo sapiens.
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524355/58.
 DR N-PSDB; AAK94181.
 XX

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2725; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX encoded by a full length human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
SQ Sequence 369 AA;

Query Match 77.3%; Score 1462; DB 22; Length 369;
Best Local Similarity 75.3%; Pred. No. 1.3e-162;
Matches 278; Conservative 44; Mismatches 40; Indels 6; Gaps 5;
QY 1 MGLRKNARNPPVLSHEPFVQNHADVMSCVGMFFVLGLMEFGTAEMSVFLLTQHGYYVVP 60
Db 1 mglrkstknppvlsqeflqnhadivscvgmffllglvfgtaeasivfltlqhsavp 60
QY 61 -AELPGSGRTLYHYGVKDLATVFFMLVAIIHATIQEYVLDKLSRRQLTKGKONKLN 119
Db 61 aaeqatgskslvygvkdlatvffmlvaliiahqeyvldklnrmqtkakqknfn 120
QY 120 EAGLSVFYIVSGTWGMIIASENCLSDPTLLWKSQPHNMFTQMKFFYSOLAYWFHSF 179
Db 121 esqgfsyvfscslwgfllsencsdptllwkarphsmntqmkffysqlaywfhaf 180
QY 180 PELYFQVRKQDIPGQIYIYGLHFGGAYLLYLNHGLMLLMLHAYVELLSVCSLLY 239
Db 181 pelyfqtktkqdlprqlvyglhfhigtayllnhglvllvhyfvelshmcglfy 240
QY 240 FGDERYQKGLSWPVPISGRVTLIVSVTVGLHLAGT-NRNGNALSGNVNVAAKIAV 298
Db 241 fsdekyqkgsllwaivflgrlvtlsvitvgfhlagsgnmpdaltgnvnvaakiav 300
QY 299 LSSCSSTQVITWTLLTVWLQWLELDANLHV-CGRKRSR-SRKGTE--GVENPNRIDS 354
Db 301 lsscsstqayvtwnltlwlqwrwvdsnlqascmkkrssrskrtengvgvetsnrvc 360
QY 355 PPRKKKEKA 362
Db 361 pprkkeks 368

RESULT 7
AAB43601
ID AAB43601 standard; Protein; 416 AA.
XX
AC AAB43601;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated protein sequence SEQ ID NO:1046.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antiadipatic; antiaesthatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.
XX Homo sapiens.
OS
PN W0200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX WPI; 2000-587533/55.
XX N-PSDB; AAC77810.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
XX useful for treating or diagnosing e.g. cancer -
XX
XX Claim 11; Page 1634-1636; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
XX in AAB43398 to AAB44239. The proteins can have activities based on the
XX tissues and cells the genes are expressed in. Example of activities
XX include: cytostatic; proliferative; vulnerable; immunomodulator;
XX antidiabetic; antiaesthatic; antirheumatic; antiarthritic;
XX antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
XX dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
XX nootropic; vasotropic; antipsoriatic and antiangiogenic. The
XX polynucleotides and polypeptides can be used for preventing, treating or
XX ameliorating medical conditions and diagnosing pathological conditions
XX polynucleotides, polypeptides, antibodies, agonists and antagonists from
XX the present invention may be used to treat immune disorders by activating
XX or inhibiting the proliferation, differentiation or mobilisation of
XX immune cells, to treat disorders of haematopoietic cells, autoimmune
XX disorders, allergic reactions, graft versus host disease and organ
XX rejection, modulate haemostatic or thrombolytic activity, modulate
XX inflammation, cancers, cardiovascular disorders, neurological disease and
XX bacterial or viral infections. The peptides, nucleotides, antibodies,
XX agonists and antagonists may be also be used in drug screens. AAC78449 to
XX AAC78457 and AAB44240 represent sequences used in the exemplification of
XX the present invention.
XX Sequence 416 AA;

Query Match 60.0%; Score 1134.5; DB 21; Length 416;
Best Local Similarity 58.7%; Pred. No. 4.5e-124;
Matches 219; Conservative 61; Mismatches 82; Indels 11; Gaps 4;
QY 1 MGLRKNARNPPVLSHEPFVQNHADVMSCVGMFFVLGLMEFGTAEMSVFLLTQHGYYVVP 60
Db 43 mairkstkspvlshefvlqnhadivscvgmffllglvfgtaeasivfltlqynvtlp 102
QY 61 A-EGLPGSGRTLYHYGVKDLATVFFMLVAIIHATIQEYVLDKLSRRQLTKGKONKLN 119
Db 103 ateeqatesvslvygvkdlatvffmlvaliiahqeyvldklnrmhsktkhskfn 162
QY 120 EAGLSVFYIVSGTWGMIIASENCLSDPTLLWKSQPHNMFTQMKFFYSOLAYWFHSF 179
Db 163 esqglsafylfacwgtfillsenysdptllwkarphsmntqmkffysqlaywfhaf 222
QY 180 PELYFQVRKQDIPGQIYIYGLHFGGAYLLYLNHGLMLLMLHAYVELLSVCSLLY 239
Db 223 pelyfqtktkqdlprqlvyglhfhigtayllnhglvllvhyfvelshisrlfy 282
QY 240 FGDERYQKGLSWPVPISGRVTLIVSVTVGLHLAGT-NRNGNALSGNVNVAAKIAV 298
Db 283 fsnekyqkgsllwaivflgrlvtlsvitvgfhlagsgnmpdaltgnvnvaakiav 342

```
QY 299 LSSSCSIQVYITWTTLTWTWQLRWLEDANLHVCGRRR-----SRSRKGTENGVE---NP 349
Db 343 lasicvtqafmmwkinfqlrrrehsafapavkkptvtkgrsskkgtengvngltls 402
QY 350 NRIDSPPKKEKA 362
Db 403 nvadspnkkks 415

RESULT 8
ID AAY48434 standard; Protein; 304 AA.
XX
AC AAY48434;
XX
DT 08-DEC-1999 (first entry)
DE Human prostate cancer-associated protein 131.
XX
KW Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
KW cancer; tissue specificity; human.
XX
OS Homo sapiens.
XX
PN DE19811194-A1.
XX
PD 16-SEP-1999.
XX
PF 10-MAR-1998; 98DE-1011194.
XX
PR 10-MAR-1998; 98DE-1011194.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosenthal A;
DR WPI; 1999-519629/44.
DR N-PSDB; AA233531.
XX
XX New nucleic acid expressed at high level in normal prostatic tissue and
PT encoded polypeptides, used to treat cancer and screen for therapeutic
PT agents.
XX
PS Claim 25; 174; 194pp; German.
XX
XX This invention describes novel nucleic acid sequences (A) that are
CC expressed at high level in normal prostatic tissue. Polypeptides (I)
CC encoded by (A) are used: (a) for identifying agents for treatment of
CC prostatic cancer and (b) for therapy of prostate cancer, optionally
CC where expressed by gene therapy methods. (A) is also used to isolate
CC full-length genes (for gene therapy) and for recombinant production of
CC (I), which can be used to raise specific antibodies. (A) are identified
CC by assembly of ESTs (expressed sequence tags) before these are analyzed
CC for expression pattern (tissue specificity). This approach eliminates
CC many of the false results, as regards tissue specificity, associated
CC with known methods that use single (usually short) ESTs. AAY48304-Y48456
CC represent peptides encoded by the expressed sequence tags described in
CC the method of the invention.
XX
SQ Sequence 304 AA;

Query Match 48.7%; Score 922; DB 20; Length 304;
Best Local Similarity 58.1%; Pred. No. 2.5e-99;
Matches 176; Conservative 48; Mismatches 69; Indels 10; Gaps 3;

QY 70 TLYHGVKDLATVFFVFMVAIIHATIQEYVDLKLRRLOLTKGKONKLNKAGQLSVFYI 129
Db 1 slyyvgirdlatvffymvaihaviqeyimdklnrmhfskthkskfnsgqlsafyl 60
QY 130 VSGIWMILTAGENCLSDPTLWKSOPHNMTFQMKFFYISOLAYWFHSPPELFOKVRK 189
Db 61 facvwtflisenyisdptllwrayphnmfkmffysqlaywlhaipelyfdtkk 120
```

```
QY 190 QDIPGQLIYIGLHDFHIGGAYLLYLNHLGLLLMLHYAVELLSSVCSLLYFGDERYQKGL 249
Db 121 ediprqlviyiglyfhiagayllnlnhglvllvlhyfveflhlsrlfyfsnekyqkgf 180
QY 250 SLWPIFVIFISGRLVTLIVSVTVVGLHLA-GTNRNGNALSGNVNVLAAKTAVLSSSCSIQVY 308
Db 181 slwavlflvlglltllsvitvgfglaraenqkldfstgnfnvlavriaviasicvtqaf 240
QY 309 ITWTTLTWTWQLRWLEDANLHVCGRRR-----SRSRKGTENGVE---NPNRIDSPPKKK 359
Db 241 mmwkinfqlrrrehsafapavkkptvtkgrsskkgtengvngltlsnvadspnkk 300
QY 360 EKA 362
Db 301 eks 303

RESULT 9
ID AAM95708 standard; Protein; 384 AA.
XX
AC AAM95708;
XX
DT 21-NOV-2001 (first entry)
DE Human reproductive system related antigen SEQ ID NO: 4366.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
```

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 03-JAN-2001; 2001US-0259678.
XX
XX

(HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-465570/50.

DR N-PSDB; AAL01678.

XX Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -

XX Claim 11; SEQ ID NO 4366; 1297pp + Sequence Listing; English.

XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.

XX Sequence 384 AA;

Query Match 39.1%; Score 739; DB 22; Length 384;

Best Local Similarity 41.8%; Pred. No. 1.le-77;

Matches 156; Conservative 62; Mismatches 135; Indels 20; Gaps 5;

QY 5 KKNARNPPVLSHEPMONHADWVSCVGMFFVLGLMFEETAEMSVLFTLQHGCVVVPBGL 64

Db 18 rrrtkysplfsqefvihnhdigfcvlcvliglmfxvtaktafllpqynisvpt--- 74

QY 65 PSGSRTL-YHYGVKDLATVFFYMLVAIIHATIQEIVLDKLSRLQLTKGKONKLNKAG 123

Db 75 -adsetvhygpkdlvtilfifitilhavvdeyildkiskrlhskvksfnesgq 133

QY 124 LSVFYIVSGIWMGITILASENCLSDPTLLWKSOPHNMTFQMKFFVISQAYWFHSFPPELY 183

Db 134 lvvfhtsvlwcfcyvvvcegyltprslwedvphvhpqvkfycqlaylhalpely 193

QY 184 FQVVRKQDIPQQLIYIGLHLFHHGAYLLYLNHLGLLMLHLYAVELSSVCSLLYFQDE 243

Db 194 fqvrkeeiprqlqyiclylvhlagayllnlsrllglllllyqstefifhcxrifyfade 253

QY 244 RYQKGLSLWPIVFSGRBLTVLIVSVTVGLHLAG-TNRNGNALSGNVNLAAKTAVLSSS 302

Db 254 nneklfsawaaxfgvtrfltlavlaigfglaromenqafdpkgnfnlcrilcvlliv 313


```
Query Match          29.9%; Score 566; DB 22; Length 368;
Best Local Similarity 37.0%; Pred. No. 2.2e-57;
Matches 136; Conservative 64; Mismatches 130; Indels 38; Gaps 9;

QY 4 RKNARNPPVLSHEFMVQNHADVMSCVGMFFVLGLMFEGETAEMSIVFLTLQHG--VWPA 61
DB 9 rKtsnknppilshfviqnhadiiscvamfvvglmnestaafasafisllhnnvsgebps 68
QY 62 EGLPSGSRITLYHYGVKDLATVFFYMLVAIIHATIOEYVLDKLSRRLQLTGKQNKLEA 121
DB 69 reqpygkptyiagikdycaiffytclimhaiqefvdkiskkhlskflarfnes 128
QY 122 GQLSVFYIVSGIWMGIILASENCLSDPTLLWKSQPHNMFTQMKFFYISQLAYWFHSPFE 181
DB 129 gqlvafyllsfvqahvllkegylgvqqlwegfdpmsflhkfyfvvqlayllhmlpe 188
QY 182 LYQKVR-KDIPGQLIYIGLHFIHGG-----AYLLYNHLGLLLMLHYAVELLSV 234
DB 189 lyfqiktkkeeqqpkivh-----sigsftliivlaytislqrlalvltlthfseillshv 242
QY 235 CSLL--YFGDERYOKGLSLWPIVIFISGRVTLIVSVTVGLHLAGTNRNGNALSQNVNL 292
DB 243 fqligvfdreerlaklrvmnnnavflirfatsvigtlyyigv-rsllaiggl--- 298
QY 293 AAKIAYLSSSCSIQVYITWTTLTWLQRLWLEDANLHVCGRKRRSRKGTENGVENPNRI 352
DB 299 -----alqgylvfiteqlrakreakk----eakreaklalktkkpktpk-- 341
QY 353 DSPPKKE 360
DB 342 dkvkrkke 349

RESULT 12
ABB67362
ID ABB67362 standard; Protein; 1575 AA.
XX
AC ABB67362;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 28878.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL11465.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 28878; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
```

```
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABJ16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1575 AA;
```

```
Query Match          29.9%; Score 566; DB 22; Length 1575;
Best Local Similarity 37.0%; Pred. No. 2e-56;
Matches 136; Conservative 64; Mismatches 130; Indels 38; Gaps 9;
```

```
QY 4 RKNARNPPVLSHEFMVQNHADVMSCVGMFFVLGLMFEGETAEMSIVFLTLQHG--VWPA 61
DB 1216 rKtsnknppilshfviqnhadiiscvamfvvglmnestaafasafisllhnnvsgebps 1275
QY 62 EGLPSGSRITLYHYGVKDLATVFFYMLVAIIHATIOEYVLDKLSRRLQLTGKQNKLEA 121
DB 1276 reqpygkptyiagikdycaiffytclimhaiqefvdkiskkhlskflarfnes 1335
QY 122 GQLSVFYIVSGIWMGIILASENCLSDPTLLWKSQPHNMFTQMKFFYISQLAYWFHSPFE 181
DB 1336 gqlvafyllsfvqahvllkegylgvqqlwegfdpmsflhkfyfvvqlayllhmlpe 1395
QY 182 LYQKVR-KDIPGQLIYIGLHFIHGG-----AYLLYNHLGLLLMLHYAVELLSV 234
DB 1396 lyfqiktkkeeqqpkivh-----sigsftliivlaytislqrlalvltlthfseillshv 1449
QY 235 CSLL--YFGDERYOKGLSLWPIVIFISGRVTLIVSVTVGLHLAGTNRNGNALSQNVNL 292
DB 1450 fqligvfdreerlaklrvmnnnavflirfatsvigtlyyigv-rsllaiggl--- 1505
QY 293 AAKIAYLSSSCSIQVYITWTTLTWLQRLWLEDANLHVCGRKRRSRKGTENGVENPNRI 352
DB 1506 -----alqgylvfiteqlrakreakk----eakreaklalktkkpktpk-- 1548
QY 353 DSPPKKE 360
DB 1549 dkvkrkke 1556
```

RESULT 13

```
AAG00189
ID AAG00189 standard; Protein; 125 AA.
XX
AC AAG00189;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 4270.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST ) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 14:57:46 ; Search time 28.82 Seconds
(without alignments)
307.651 Million cell updates/sec

Title: US-09-807-470-2
Perfect score: 1892
Sequence: 1 MGLRKNARNPVLSHEFMV.....NGVENPNRIDPPKKKEKAP 363

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA: *
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep: *
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep: *
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep: *
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep: *
5: /cgn2_6/ptodata/2/iaa/PTDUS_COMB.pep: *
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	148.5	7.8	411	2	US-08-336-031-2
2	148.5	7.8	411	2	US-08-902-853-7
3	148.5	7.8	411	5	PCT-US95-06725-2
4	138.5	7.3	394	2	US-08-902-853-1
5	125	6.6	387	2	US-08-902-853-6
6	110	5.8	380	2	US-08-902-853-3
7	89	4.7	649	4	US-08-800-291B-5
8	89	4.7	649	4	US-08-800-291B-6
9	89	4.7	650	4	US-08-800-291B-4
10	87	4.6	1476	4	US-09-256-703-2
11	87	4.6	1479	2	US-08-951-912-4
12	87	4.6	1479	4	US-09-174-077-4
13	87	4.6	1480	1	US-07-637-621-2
14	87	4.6	1480	1	US-08-136-742A-2
15	87	4.6	1480	1	US-08-135-809A-2
16	87	4.6	1480	1	US-08-466-886-17
17	87	4.6	1480	1	US-08-951-912-2
18	87	4.6	1480	2	US-08-951-912-6
19	87	4.6	1480	2	US-08-469-461-2
20	87	4.6	1480	2	US-08-469-461-4
21	87	4.6	1480	2	US-08-691-605-2
22	87	4.6	1480	2	US-08-455-552A-14
23	87	4.6	1480	3	US-07-890-609-2
24	87	4.6	1480	3	US-07-890-609-4
25	87	4.6	1480	3	US-09-248-026-2
26	87	4.6	1480	4	US-08-469-617-17
27	87	4.6	1480	4	US-08-681-838A-2

28	87	4.6	1480	4	US-08-681-838A-3	Sequence 3, Appli
29	87	4.6	1480	4	US-09-174-077-2	Sequence 2, Appli
30	87	4.6	1480	4	US-09-174-077-6	Sequence 6, Appli
31	87	4.6	1480	5	PCT-US93-11667-2	Sequence 2, Appli
32	87	4.6	1480	6	5240846-5	Patent No. 5240846
33	85	4.5	193	2	US-08-336-031-6	Sequence 6, Appli
34	85	4.5	193	5	PCT-US95-06725-6	Sequence 6, Appli
35	85	4.5	2183	3	US-08-746-111-5	Sequence 5, Appli
36	80	4.2	617	1	US-07-879-617A-11	Sequence 11, Appli
37	80	4.2	617	1	US-08-301-722A-3	Sequence 3, Appli
38	80	4.2	617	1	US-08-240-783B-3	Sequence 3, Appli
39	80	4.2	617	1	US-08-753-985-11	Sequence 11, Appli
40	80	4.2	617	3	US-09-084-813-3	Sequence 3, Appli
41	80	4.2	617	5	PCT-US92-09662-3	Sequence 3, Appli
42	79	4.2	326	4	US-08-986-768-2	Sequence 2, Appli
43	79	4.2	602	1	US-08-295-814E-2	Sequence 2, Appli
44	79	4.2	602	4	US-09-343-361-2	Sequence 2, Appli
45	79	4.2	602	5	PCT-US93-01959-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-336-031-2
; Sequence 2, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EURARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 9303Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-031-2

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 1.1e-08;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MYSCVGMFFVLGLMFGTAEISMIVFLTLQGVVVVPAEGLPSGSRITLYHYGVKDLATVFFV 85

Db 91 LVCVTSAYFLSGNRRTESNPLHMFVAISYQ-----VDGTDS-----YAKGKDLFSVFY 139
QY 86 MVAIIIIHATIQEYVLDKLSRLQITKQKONKLEAGOLSVFYI-VSGIWMIL-ASEN 143
Db 140 MIFTFLEFLMDVVIRPFTYVLTNSVSHRQKRMLEQMYAIFYCGVSPFGLIYIMYHSDL 199
QY 144 CLSDPTLLWKSOPHNMTFMKFFYISQAYW-----FHSFPE 181
Db 200 WLFKTKPMYRTPVITNPFLEKIFVLGQAAFWAQACVLVLQLEKPRKDYKELVFHHVT 259
QY 182 L-----YFQVKRQDIPGQIYI-----GLHLFH 205
Db 260 LLLIWSVVFHFTKM-----GLAIYITMDVSDFFLSKLTNLYNSVFTPFVGLFVF- 312
QY 206 IGGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDERYOKGLSLWPIVFSIGRLVTLI 265
Db 313 ---FWIYLRHVNVNIRILWSVLTFRHGNVYLNFAQOQYKWCISL-PIVFV-----LI 361
QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370

RESULT 2

US-08-902-853-7
; Sequence 7, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 541568
; US-08-902-853-7

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 1.le-08;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MVSQGVMEFVLGMPFGTAEMSIVFLTLQHGWWVPAEGLPSGSRITLYHYGVKDLATVFFY 85
Db 91 LVCVTSAYFLSGNRRTESNPLHMFVAISYQ-----VDGTDS-----YAKGKDLFSVFY 139
QY 86 MVAIIIIHATIQEYVLDKLSRLQITKQKONKLEAGOLSVFYI-VSGIWMIL-ASEN 143
Db 140 MIFTFLEFLMDVVIRPFTYVLTNSVSHRQKRMLEQMYAIFYCGVSPFGLIYIMYHSDL 199
QY 144 CLSDPTLLWKSOPHNMTFMKFFYISQAYW-----FHSFPE 181
Db 200 WLFKTKPMYRTPVITNPFLEKIFVLGQAAFWAQACVLVLQLEKPRKDYKELVFHHVT 259
QY 182 L-----YFQVKRQDIPGQIYI-----GLHLFH 205
Db 260 LLLIWSVVFHFTKM-----GLAIYITMDVSDFFLSKLTNLYNSVFTPFVGLFVF- 312
QY 206 IGGAYLLYLNHLGLLLMLHYAVELLSVCSLLYFGDERYOKGLSLWPIVFSIGRLVTLI 265
Db 313 ---FWIYLRHVNVNIRILWSVLTFRHGNVYLNFAQOQYKWCISL-PIVFV-----LI 361
QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370

RESULT 3

PCT-US95-06725-2
; Sequence 2, Application PC/TUS9506725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EUKARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06725
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875 & 08/336,031
; FILING DATE: 03-JUN-1994 & 08-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06725-2


```

Query Match          6.6%; Score 125; DB 2; Length 387;
Best Local Similarity 23.9%; Pred. No. 5.4e-06;
Matches 57; Conservative 35; Mismatches 100; Indels 46; Gaps 11;

QY      65   PGGSRITYHYGVKDLATVFYFMYMLVAIIHHATHIOEYVLDKLSRRQLTKGQN-----KLN 119
        |||  |||  |||  |||  |||  :|||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      104   PDGS---YKGPRKDACFPFWIVTFATFRVIMDYVE---RPFVLNMGVRNRKVIRFC 156

QY      120   EAQGLSVEYIVSGIWGMILLASENCSDPTLLNKSOHPNNMWTQMKEFYISOLAYWFHSF 179
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      157   EQSYSPFYILCFWELGLIYYRSNYWSNEEKLEDPQQYMSPLFKRAYILIQLGFWLQOI 216

QY      180   PELYFOVKRKQD-----IPGQLIYI--GLHLPFHIGAYILYLNLHLGLLLL----ML 224
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      217   LVLHLSEQ-RRADHWQMPFAHIIVTCALLILSYGNFLRVGNA-ILYIFDLSDYILSGGKML 274

QY      225   HYAEVLLSSVCSLLYFGDERYQGSLSWPIVFISGRLVTLIVSVTV-----GLHL 275
        |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      275   KYLG--FGKICDYLF-----GIFVASWYSRHLYFSKILRVVVVNAPEIIGGFHL 322


RESULT      6
US-08-902-853-3
; Sequence 3, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2NOT01
; CLONE: 493014
; US-08-902-853-3

```

Db 271 VFFSCVISLYHGLMQWILKIAWLMQVTGTATETLSVAGNIEVSQTEAP-----L 324
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFVIS-----QLAY 174
Db 325 LIRPYLADMTL---SEVHVMTGGYATAGSLGAYISFGIDATSLIAASVMAAPCALAL 381
QY 175 WFHSFPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNHLGLLL---LMLHYAVELL 231
Db 382 SKLIVPEVESFRREE-----GVKLTYGDAQNLEIAASTGAISVKVAVANIANLI 433
QY 232 SVV-----CSLLYFGDERYQKGLS-----LWPIVIFISG 259
Db 434 AFLAVLDFINAALSGLMDWDIOGLSFQGLCSYLIRPVAFLMG 476

RESULT 8

US-08-800-291B-6
; Sequence 6, Application US/08800291B
; Patent No. 6153740
; GENERAL INFORMATION:
; APPLICANT: J.D. Young & C.E. Cass
; TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,291B
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/499,314
; FILING DATE: 7-JULY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/044WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 649 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-291B-6

Query Match 4.7%; Score 89; DB 4; Length 649;
Best Local Similarity 21.9%; Pred. No. 0.2;
Matches 62; Conservative 49; Mismatches 102; Indels 70; Gaps 11;
QY 30 VGMFFVLGMFEGTAEMSIVFTLQHGVVVPAEGLPSGSRTHYHG-VKDL-----AT 81
Db 211 LGLQFVLGLLVIRTEPGTAFEGWILKIAWLMQVTGTATETLSVAGNIEVSQTEAP-----L 324
QY 82 VFYFVLVAIIHATIOEYVLDKLSRRLQTLTKG--KQKLINEAGQLSVFIVSGWMIIL 139
Db 271 VFFSCVISLYHGLMQWILKIAWLMQVTGTATETLSVAGNIEVSQTEAP-----L 324
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFVIS-----QLAY 174

Db 325 LIRPYLADMTL---SEVHVMTGGYATAGSLGAYISFGIDATSLIAASVMAAPCALAL 381
QY 175 WFHSFPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNHLGLLL---LMLHYAVELL 231
Db 382 SKLIVPEVESFRREE-----GVKLTYGDAQNLEIAASTGAISVKVAVANIANLI 433
QY 232 SVV-----CSLLYFGDERYQKGLS-----LWPIVIFISG 259
Db 434 AFLAVLDFINAALSGLMDWDIOGLSFQGLCSYLIRPVAFLMG 476

RESULT 9

US-08-800-291B-4
; Sequence 4, Application US/08800291B
; Patent No. 6153740
; GENERAL INFORMATION:
; APPLICANT: J.D. Young & C.E. Cass
; TITLE OF INVENTION: CDNA ENCODING NUCLEOSIDE TRANSPORTER
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,291B
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/499,314
; FILING DATE: 7-JULY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07254/044WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-800-291B-4

Query Match 4.7%; Score 89; DB 4; Length 650;
Best Local Similarity 21.9%; Pred. No. 0.2;
Matches 62; Conservative 49; Mismatches 102; Indels 70; Gaps 11;
QY 30 VGMFFVLGMFEGTAEMSIVFTLQHGVVVPAEGLPSGSRTHYHG-VKDL-----AT 81
Db 212 LGLQFVLGLLVIRTEPGTAFEGWILKIAWLMQVTGTATETLSVAGNIEVSQTEAP-----L 271
QY 82 VFYFVLVAIIHATIOEYVLDKLSRRLQTLTKG--KQKLINEAGQLSVFIVSGWMIIL 139
Db 272 VFFSCVISLYHGLMQWILKIAWLMQVTGTATETLSVAGNIEVSQTEAP-----L 325
QY 140 ASENCLSDPTLLWKSPHNMNT-----FQMKFFVIS-----QLAY 174
Db 326 LIRPYLADMTL---SEVHVMTGGYATAGSLGAYISFGIDATSLIAASVMAAPCALAL 382
QY 175 WFHSFPELYFQKVRKODIPGQLIYIGLHLFHIGGAYLLYLNHLGLLL---LMLHYAVELL 231
Db 383 SKLIVPEVESFRREE-----GVKLTYGDAQNLEIAASTGAISVKVAVANIANLI 434

```

QY 232 SSV-----CSLLYFGDERYOKGLS-----LWPIVIFISG 259
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 AFLAVLDFINALSALWGLMDVLDIQLSFLQICSYIIRPVAFILMG 477
; : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US-09-256-703-2
; Sequence 2, Application US/09256703
; Patent No. 6294379
; GENERAL INFORMATION:
; APPLICANT: Dong, Jian-yun
; APPLICANT: Kan, Yuet Wai
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Efficient AAV Vectors
; FILE REFERENCE: 023070-084910US
; CURRENT APPLICATION NUMBER: US/09/256,703
; FILING DATE: 1999-02-24
; PRIOR FILING DATE: 1998-02-25
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1476
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: truncated cystic fibrosis transmembrane
US-09-256-703-2

```

```

Query Match 4.6%; Score 87; DB 4; Length 1476;
Best Local Similarity 20.8%; Pred. No. 1.2;
Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;

```

```

QY 24 ADMVSCVGMFFVLGIMFEGTAEMSIIVFLTLQHGCVVPAEGLPSGRTLYHYGV-----KD 78
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 923 ADTLAMGFFRGPLVHTLITVSKILHKLHLSVL---QAPMSTLNTLTKAGGLNRFSD 979
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 LA-----TVF-FYMLVAIIH-----ATIOEYVLDK-----LSRRLQLTGKQ 115
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 980 IAILDDLPLTIFDFIQLLLIVIGAVAVLQPIFVATVPVIVAFIMLRAYFIQTSSQ 1039
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 NK-LNEAGQLSVF-YIVSGIWMIIASENCISDPTLLWKSQPHNMFTOMKFFYISOL- 172
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1040 LKQLESGRSPFTHLVTSKLGWTLRA-----FGRQPY-----FETLFHKNLH 1085
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 -AYWFHSPPELYFQVKRKQDIPGQIYIGLHFIHGAYLLYL-----ERYOKGL- 249
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1086 TANWF-----LYLSTLRWFQMRIMFV---IFFIAVTFISILTTGEGEGRVGIILTLAM 1137
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 -----LHYAV-----ELLSSVCSLLYFGD-----ERYOKGL- 249
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1138 NIMSTLQAVNSIDVDSLMSRSVRVFKFIDMPTGKPTKTRPKYKNGQLSKVMIIENSH 1197
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 -----SLWPIVIFISGRVLT--IVSVTVVGLHLAGTNRNGAL----- 285
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1198 VKKDDIWP-----SGQMTVKDLTKAYTEG-----GNAILENISFISIPGQVRVGLLG 1244
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 ----SGNVNVLAAKIAVLSSSCSIQV-YITWTLTTVWLQW 321
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1245 RTGSGKSTLLSAFLRLNTEGEIQIDGVSW--DSITLQW 1282
; : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 11
US-08-951-912-4
; Sequence 4, Application US/08951912
; Patent No. 5972995
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC
; TITLE OF INVENTION: FIBROSIS THERAPY
; NUMBER OF SEQUENCES: 6

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,912
; FILING DATE: 16-OCT-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 200116.403
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1479 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-951-912-4

```

```

Query Match 4.6%; Score 87; DB 2; Length 1479;
Best Local Similarity 20.8%; Pred. No. 1.3;
Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;

```

```

QY 24 ADMVSCVGMFFVLGIMFEGTAEMSIIVFLTLQHGCVVPAEGLPSGRTLYHYGV-----KD 78
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 922 ADTLAMGFFRGPLVHTLITVSKILHKLHLSVL---QAPMSTLNTLTKAGGLNRFSD 978
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 79 LA-----TVF-FYMLVAIIH-----ATIOEYVLDK-----LSRRLQLTGKQ 115
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 979 IAILDDLPLTIFDFIQLLLIVIGAVAVLQPIFVATVPVIVAFIMLRAYFIQTSSQ 1038
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 NK-LNEAGQLSVF-YIVSGIWMIIASENCISDPTLLWKSQPHNMFTOMKFFYISOL- 172
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1039 LKQLESGRSPFTHLVTSKLGWTLRA-----FGRQPY-----FETLFHKNLH 1084
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 173 -AYWFHSPPELYFQVKRKQDIPGQIYIGLHFIHGAYLLYL-----ERYOKGL- 249
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1085 TANWF-----LYLSTLRWFQMRIMFV---IFFIAVTFISILTTGEGEGRVGIILTLAM 1136
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 -----LHYAV-----ELLSSVCSLLYFGD-----ERYOKGL- 249
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1137 NIMSTLQAVNSIDVDSLMSRSVRVFKFIDMPTGKPTKTRPKYKNGQLSKVMIIENSH 1196
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 250 -----SLWPIVIFISGRVLT--IVSVTVVGLHLAGTNRNGAL----- 285
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1197 VKKDDIWP-----SGQMTVKDLTKAYTEG-----GNAILENISFISIPGQVRVGLLG 1243
; : : : : : : : : : : : : : : : : : : : : : : : : :
QY 286 ----SGNVNVLAAKIAVLSSSCSIQV-YITWTLTTVWLQW 321
; : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1244 RTGSGKSTLLSAFLRLNTEGEIQIDGVSW--DSITLQW 1281
; : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
US-09-174-077-4
; Sequence 4, Application US/09174077
; Patent No. 6329422
; GENERAL INFORMATION:
; APPLICANT: Fischer, Horst
; APPLICANT: Illek, Beate
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CYSTIC FIBROSIS THERAPY

```



```

; FILE REFERENCE: 200116.403C1
; CURRENT APPLICATION NUMBER: US/09/174,077
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: US 08/951,912
; EARLIER FILING DATE: 1997-10-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1479
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-174-077-4

```

Query Match 4.6%; Score 87; DB 4; Length 1479;
Best Local Similarity 20.8%; Pred. No. 1.3;
Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;

Qy	24	ADWVCVGMPFVLGLMFECTAEMSIVFLQHGVVVVPAEGLPSGSRITLYHHYGV-----KD	78
Dd	922	ADTLAMGEFRGLPVLHTLITVYSKILHHKMHLSSL---QAPMSTLNTLKAGILNRFSSQ	978
Qy	79	LA-----TVP-FYMLVAIIH-----ATIQEVLDK-----LSRRLQLTGKGQ	115
Dd	979	IAILDDLPLUTFDFOQLGLVIGAIIVAVLQPIFVATVPVIVAFIMLAYLOTSQQ	1038
Qy	116	NK-LNEAGOLSVF-YIVSGIWGMILASBNCLSDPTLLWKSQPHNMFTQMFKFYISOL-	172
Dd	1039	LKOLESEGRSPITHLVTSLGWLTLRA-----FGROPY---FEFLFHKALNLH	1084
Qy	173	-AYWFHSFPFLYFQVKRKODIPQLIYYIGLHLFHTGGAYLIYL-----NHGLLLLLM--	223
Dd	1085	TANWF-----LYLSTLRWFOMRIEFV---IFFIAVTFSILTAGEGRVGIIILIAM	1136
Qy	224	----LHYAV-----ELLSSVCSLLFGD-----ERXQKGL-----	249
Dd	1137	NIMSTQWAVNSIDVDSLMRSVSRVFIDMPTBEGTKTKPKYKNQSLSKVMIIENSH	1196
Qy	250	----SLWPVFVIFSGRLVTL--TVSVVTVVGLHLAGTNRRGNAL-----	285
Dd	1197	VKKDDIWP---SGQOMTKDLTAKYTEG-----GNILENISFISPGORVGLLG	1243
Qy	286	----SGNVNVLAKIAVLSSCSCIQV-YITWTLTVTVMLOWR	321
Dd	1244	RTSGSKSTLSAFLRLNTEGIIQIDGYSW--DSTITLOW	1281

```

RESULT 13
US-07-637-621-2
; Sequence 2, Application US/07637621
; Patent No. 5407796
; GENERAL INFORMATION:
; APPLICANT: cutting, gary
; APPLICANT: antonarakis, stylianos e
; APPLICANT: kazazian jr., haig h
; TITLE OF INVENTION: CYSTIC FIBROSIS MUTATION CLUSTER
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie and Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07637,621
; FILING DATE: 19910104
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

NAME: kagan, sarah a
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.030010
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: HOMO SAPIENS
US-07-637-621-2

Query Match	4.6%	Score 87;	DB 1;	Length 1480;
Best Local Similarity	20.8%	Pred. No. 1.3;		

Qy	24	ADMVSCVGMFFVLGLMEFGTAEMSIVFLQLHGVVVPABGLPESGRTLYHHGV-----KD	78
Dd		: : : : : : : : : : : : : : :	
Db	923	ADTLAMGGFRGLPLVHTLITVSKILHKKMLHSVL---QAPMSTLNTLKAGILNRFSQD	979
Qy	79	LA-----TVF-FYMLVAIIIH-----ATIQEYVLDK-----LSRLLOLTGKGQ	115
Dd		: : : : : : : : : : :	
Db	980	TAILDLLPTIFEDTOLLIVIGAIANVAVLQPIFYVATVPVIVAFIMLRAYLQTSSQD	1039
Qy	116	NK-LNEAQSLSVF-YIVSGIWGMIIIASENCSDPTLLWKSPHNMMTFQMKFFVISOL-	172
Dd		- - - - - : : : : : : : : : : : : :	
Db	1040	LKQLESEGRSPIHLVTSIKGLWTLRA-----FGRQPY---FETLFKHALNLH	1085
Qy	173	-AYWFHSPEPEYEQVKVRKODIPQLIYIGLHFHGGAYLLYL-----NHLGLLLL--	223
Dd		: : : : : : : : : : : : : : :	
Db	1086	TANWF-----LYLSTLRWQMRIEMIFV---IFFIAVTISILTGEGERVGIIILTILAM	1137
Qy	224	----LHYAV-----ELLSSVCSSLYPGD-----ERYQKGL-----	249
Dd		: : : : : : : : : : :	
Db	1138	NIMSTLOWANSSIDVDLSMRSVSRFKRIDMPTEGKPTKTKPKYGKQLSKVMIIENSH	1197
Qy	250	----SLWPPIVIFSGRLVTL-IVSVVTVGLHLAGTNRGNAL-----	285
Dd		: :	
Db	1198	VKDDIWP-----SGGQMTVKDLTAKYTEG-----GNAILENISFISPGORVGLLG	1244
Qy	286	----SGNVNVLAAKI AVLSSCSIQV-YITWTLTTVWLQRW	321
Dd		: : : : : : : : : : : : : : :	
Db	1245	RTGSKASTLLSAPRLNTEGEI QIDGVSW---DSITLQOW	1282

```

RESULT 14
US-08-136-742A-2
; Sequence 2, Application US/08136742A
; Patent No. 5670488
; GENERAL INFORMATION:
; APPLICANT: Gregory, R.J., Armentano, D., Couture, L.,
; APPLICANT: A.E.
; TITLE OF INVENTION: GENE THERAPY FOR CYSTIC FIBROSIS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BRUMBAUGH, GRAVES, DONOHUE & RAYMOND
; STREET: 30 ROCKEFELLER PLAZA
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/136-742A

```

;; FILING DATE: 02-DEC-1993
;; CLASSIFICATION: 514
;; PRIOR APPLICATION NUMBER: 514
;; APPLICATION DATA: US 07/985,478
;; FILING DATE: 02-DEC-1992
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seide, Rochelle K.
;; REGISTRATION NUMBER: 32,300
;; REFERENCE/DOCKET NUMBER: A30668 (Genzyme Dkt. IG4-9.11)
;; TELEPHONE: (212) 408-2500
;; TELEFAX: (212) 763-2519
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1480 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-136-742A-2

COUNTRY: USA
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/135,809A
FILING DATE: 13-OCT-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: LASSEN, ELIZABETH
REGISTRATION NUMBER: 31,845
REFERENCE/DOCKET NUMBER: IG4-9.12
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1480 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-135-809A-2

Query Match
Best Local Similarity 4.6%; Score 87; DB 1; Length 1480;
Matches 83; Conservative 63; Mismatches 112; Indels 142; Gaps 24;

QY 24 ADMVSCVGMFFVLGLMFEGETAEMSIIVLTLOHGVVPAEGLPSGSRTRYHYGV-----KD 78
Db 923 ADTLAMGFFRGLPLVHTLTVSKILHKLHLSVL---QAPMSTLNTLKAGGILNRFSD 979
QY 79 LA-----TVP-FYMLVAIIH-----ATIQEYVLDK-----LSRRLQITKQK 115
Db 980 IAILDDLLPTIFDFIOLLIVIGATAVAVLQPIYFVATVPVIVAFIMLRAYFLQTSQ 1039
QY 116 NK-LNEAGQLSVF-YIVSGIWGMIIILASENCSDPTLLWKSQPHNMFTQMKFFYISOL- 172
Db 1040 LKOLESGRSPITHLVTSKGLWTLRA-----FGRQPY-----FETLFHKAALNH 1085
QY 173 -AYWFSHFPELYFQVKRQDIPQOLIYIGLHFLHIGGAYLLYL-----NHLGLLLM-- 223
Db 1086 TANWF-----LYLSTURWFQRMIEFV---IFFIAVTFISLTGTGEGRVIILTLAM 1137
QY 224 -----LHYAV-----ELLSSVCSLLYFGD-----ERYQGL----- 249
Db 1138 NIMSTLOWAVNSSIDVSLMRSVSRVFKFIDMPTGKPTKSTPKYKNGOLSKVMIIENSH 1197
QY 250 ----SLWPVIFSGRLVTL--IVSVTVVGLHLAGTNRGNAL----- 285
Db 1198 VKKDDIWP-----SGQMTVKDLTKAYTEG-----GNAILNIFSISPCQRVGLLG 1244
QY 286 ---SGNVNVLAAKIAVLSSSCSIQV-YITWTLTVMLOWR 321
Db 1245 RTGSGKSTLLSAFLRLNTEGEIQIDGVSW--DSITLQOW 1282

RESULT 15
US-08-135-809A-2
; Sequence 2, Application US/08135809A
; Patent No. 5688677
; GENERAL INFORMATION:
; APPLICANT: CHENG, SENG H.
; APPLICANT: DITULLIO, PAUL
; APPLICANT: EBERT, KARL M.
; APPLICANT: MEADE, HARRY M.
; APPLICANT: SMITH, ALAN E.
; TITLE OF INVENTION: DROXYRIBONUCLEIC ACIDS CONTAINING
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS

Search completed: September 6, 2002, 17:10:38
Job time: 7972 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 15:18:41 ; Search time 52.02 seconds
(without alignments)
670.520 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKKNARNPPVLSHFPMV.....NGVENPNRIDSPPKKKKAP 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1150.5	60.8	374	2	S21736	translocating chai
2	1134.5	60.0	374	2	S30034	translocating chai
3	441.5	23.3	371	2	T19417	hypothetical prote
4	434.5	23.0	373	2	T19419	hypothetical prote
5	148.5	7.8	411	2	S46800	LAG1 protein - yea
6	137	7.2	418	2	S30134	hypothetical prote
7	133	7.0	357	2	T40389	longevity assuranc
8	125	6.6	390	2	T38012	longevity-assuranc
9	116.5	6.2	696	2	B86726	hypothetical prote
10	107.5	5.7	614	2	A98845	Na+/H+ antiporter
11	103.5	5.5	393	2	E96954	Na/H antiporter (n
12	102.5	5.4	522	2	B83987	ABC transporter (p
13	101	5.3	308	2	H86268	hypothetical prote
14	100.5	5.3	360	2	T27324	hypothetical prote
15	100.5	5.3	397	2	D71467	probable tyrosine
16	99.5	5.3	395	2	D81729	Mtr/TnaB/Tyros perm
17	99	5.2	286	2	AG0403	anaerobic dimethyl
18	98.5	5.2	722	2	G83685	hypothetical prote
19	97.5	5.2	397	2	T00098	hypothetical prote
20	97	5.1	372	2	AH0703	probable membrane
21	96.5	5.1	534	2	D71698	cytochrome-c oxida
22	95.5	5.0	333	2	AI0050	probable ABC trans
23	95.5	5.0	411	2	G90154	arsenite transport
24	95	5.0	370	2	H64926	probable membrane
25	95	5.0	370	2	C90528	hypothetical prote
26	95	5.0	370	2	G85776	hypothetical prote
27	95	5.0	509	2	G81929	probable iron-upta
28	94.5	5.0	540	1	I49454	sterol O-acetyltrans
29	94.5	5.0	547	2	T27253	hypothetical prote

30 94.5 5.0 788 2 S48191 probable ubiquinol
31 94 5.0 707 2 T09340 hypothetical prote
32 94 5.0 1224 2 H96615 hypothetical prote
33 94 5.0 1450 2 JC6139 cystic fibrosis tr
34 93.5 4.9 519 2 D85437 hypothetical prote
35 93.5 4.9 521 2 A32431 cytochrome-c oxida
36 93 4.9 417 2 AB0682 probable membrane
37 93 4.9 495 2 E81251 NADH dehydrogenase
38 93 4.9 515 2 B90504 amino acid transpo
39 93 4.9 531 2 T40575 major facilitator
40 93 4.9 532 1 E69821 multidrug resistanc
41 93 4.9 542 2 E90604 hypothetical prote
42 92.5 4.9 353 2 AF0676 hydrogenase-1 oper
43 92.5 4.9 723 2 T21869 hypothetical prote
44 92 4.9 345 2 AC1252 conserved hypothet
45 92 4.9 447 2 D90012 hypothetical prote

ALIGNMENTS

RESULT 1

S21736

translocating chain-associating membrane protein - dog

C;Species: Canis lupus familiaris (dog)

C;Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1999

C;Accession: S21736

R;Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.

Nature 357, 47-52, 1992

A;Title: A protein of the endoplasmic reticulum involved early in polypeptide translo

A;Reference number: S21736; MUID:92244357

A;Accession: S21736

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-374 <GOE>

A;Cross-references: EMBL:X63678; NID:g941; PIDN:CAA45217.1; PID:g942

C;Superfamily: translocating chain-associating membrane protein

C;Keywords: transmembrane protein

Query Match 60.8%; Score 1150.5; DB 2; Length 374;
Best Local Similarity 59.8%; Pred. No. 1.8e-92;
Matches 223; Conservative 58; Mismatches 81; Indels 11; Gaps 4;

QY 1 MGLRKKNARNPPVLSHFPMVONHADVMSCVGMFFVLGLMFEGETAEMSIIVLTLOHGVVVP 60

Db 1 MAIRKSTKSPVLSHEFILQNHADIVSCVAVFLLGLMFEITAKASIIIVTLQNTVLP 60

QY 61 A-EGLPSSRTLYHYGVKDLATVFFYMLVAIIIIHATIQEYVLDKLSRLQLTKGKQKLN 119

Db 61 ATEQOATESLXYGKIDLATVFFYMLVAIIIIHATIQEYVLDKLSRLQLTKGKQKLN 120

QY 120 EAGLSVFIYVSGTWGMIIASENCLSDPTLLNKSQPINMTFQMKFYISOLAYWFHSF 179

Db 121 ESGQLSAFYLFSCIWGTFILISYISDPTILWRAYPHNMTFQMKFFYIAQLAYWFHAF 180

QY 180 PELYFQVKRQDIPQLYIGLHLFHIGVALLVNLHGLLLHLHVAVELLSVCSLLY 239

Db 181 PELYFQVKRQDIPQLYIGLHLFHIGVALLVNLHGLLLHLHVAVELLSVCSLLY 240

QY 240 FGDERYQKGLSLWPIVFTSGRLVTLIVSVVTVGLHLA-GTNRNGNALSGNVNLAALIAV 298

Db 241 FSDEKYQKGLSLWPIVFTSGRLVTLIVSVVTVGLHLA-GTNRNGNALSGNVNLAALIAV 300

QY 299 LSSSCSIQVYITWTLTTVWLQRLWLEDANLHVCGRKRR-----SRSRKKTENGVE---NP 349

Db 301 LASICITQAFMMWKFNFQLRWRREHSTFOAPVVKKPTVTKGRSRKKTENGVTGTVTS 360

QY 350 NRIDSPPKKKEKA 362

Db 361 NGADSPRNKKEKS 373

RESULT 2

S30034
translocating chain-associating membrane protein - human
C:Species: Homo sapiens (man)
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C:Accession: S30034
R:Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.
A:Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A:Reference number: S21736; MUID:92244357
A:Accession: S30034
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <GOE>
A:Cross-references: EMBL:X63769; NID:g37264; PIDN:CAA45218.1; PID:g37265
C:Superfamily: translocating chain-associating membrane protein

Query Match 60.0%; Score 1134.5; DB 2; Length 374;
Best Local Similarity 58.7%; Pred. No. 4.6e-91;
Matches 219; Conservative 61; Mismatches 82; Indels 11; Gaps 4;
QY 1 MGLKKARNPPVLSHEFMVONHADWYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 60
Db 1 MAIRKKSTKPPVLSHEFVLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNNVTL 60
QY 61 A-EGLPGSRTLYHYGVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLQTLTKGKONKLN 119
Db 61 ATERQATESVSLYYIGIKDLATVFFVLMVAIIHATIQEYVLDKLSRRLQTLTKGKONKLN 120
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMTFQMKFFYISQLAYWFHSF 179
Db 121 ESGLSAFYLFACVWGTFILISENYISDPTILWRAPYHNLMTFQMKFFYISQLAYWLF 180
QY 180 PELYFQVRKQDIPGQLIYIGLHFGGAYLLYLNHLGLLMLHYAVELLSSVCLLY 239
Db 181 PELYFQTKKEDIPQLYVIGLHFGAGAYLLYLNHLGLLMLHYAVELLSSVCLLY 240
QY 240 FGBERTQKLSLMPVIFISGRVLTIVSVVTVGLHLA-GTNRNGNALSGNVNVAIAKIAV 298
Db 241 FSNERYKQKSLWAVLVGLRLTLILSVLTVGFLARAENQKLDFTSGNENVAIRIAV 300
QY 299 LSSSCSIQVYITWTLTVWLRWLEDANLHVCGRKR-----SRSRKKTENGVE---NP 349
Db 301 LASICVTQAFMWMKFINQLRRHSAFAQPAVKKKPTVTKGRSSKKGTVNGVNGTLTS 360
QY 350 NRIDSPPKKKEA 362
Db 361 NVADSPRNKKEKS 373

RESULT 3

T19417
hypothetical protein C24F3.la - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19417
R:McMurray, A.
A:Reference number: Z19122
A:Accession: T19417
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <WIL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18770.1; GSPDB:GN00022; CESP:C24F3.la
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.la
A:Map position: 4
A:Introns: 114/3; 158/2; 363/3
C:Superfamily: translocating chain-associating membrane protein

Query Match 23.3%; Score 441.5; DB 2; Length 371;
Best Local Similarity 30.0%; Pred. No. 9.8e-31;
Matches 112; Conservative 76; Mismatches 152; Indels 33; Gaps 11;
QY 2 GLRKKARNPPVLSHEFMVONHADWYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 61
Db 7 GSKASKKPPVLSHEFVLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNNVTL 66
QY 62 EGLPGSRTLYHYGVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLQTLTKGKONKLN 119
Db 67 AVEQGOEREVHGVLGILDLPAIFFYVSVCIWVHAVVQEVGLDKISKKTHLSKVSTFKFG 126
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMTFQMKFFYISQLAYW 175
Db 127 ESPH-QMFTVYISIAHAFYIVSERLEDFSEKSWLGYPTHEHRVMSAAAYKLYFIQISY 185
QY 176 FHSFPELYFQVRKQDIPGQLIYIGLHFGGAYLLYLNHLGLLMLHYAVELLSSV 235
Db 186 IHQFPFYLQKLRDEIRKQSVQAILHAFISYAFNFRVGLALITLEYITQLIFHIA 245
QY 236 SLLYFGDERYQKGLS-----LWPIFVIGRLVTLIVSVTV--VGLHLAGTNRNGNALSGN 288
Db 246 RFAHFVG---RKGLSDPAFKLNGSPVLVRLGSIILIAVMTFWYGLRQA-ESPVDISAGN 301
QY 289 VNVLAAKIAVLSSCSIQVYITWTLTVWLRWLEDANLHVCGRKRKRSGKTENGVEN 348
Db 302 FNTAVIRLNVLLAVLQLFLLYSFVV-----FHM-GRFESNAKREKKKSA 349
QY 349 PNRIIDSPPKKKEK 361
Db 350 AAAY---PKKEK 359

RESULT 4

T19419
hypothetical protein C24F3.la - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T19419
R:McMurray, A.
A:Reference number: Z19122
A:Accession: T19419
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-373 <WIL>
A:Cross-references: EMBL:AL022716; PIDN:CAA18772.1; GSPDB:GN00022; CESP:C24F3.la
A:Experimental source: clone C24F3
C:Genetics:
A:Gene: CESP:C24F3.la
A:Map position: 4
A:Introns: 114/3; 158/2; 365/3
C:Superfamily: translocating chain-associating membrane protein

Query Match

Best Local Similarity 23.0%; Score 434.5; DB 2; Length 373;
Matches 113; Conservative 75; Mismatches 151; Indels 37; Gaps 12;
QY 2 GLRKKARNPPVLSHEFMVONHADWYSCVGMFFVLGLMFEGETAEMSIIVLTLOHGYYVVP 61
Db 7 GSKASKKPPVLSHEFVLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNNVTL 66
QY 62 EGLPGSRTLYHYGVKDLATVFFVLMVAIIHATIQEYVLDKLSRRLQTLTKGKONKLN 119
Db 67 AVEQGOEREVHGVLGILDLPAIFFYVSVCIWVHAVVQEVGLDKISKKTHLSKVSTFKFG 126
QY 120 EAGLSVFIYVSGIWMIIASENCLSDPTLLWKSSOPHNMTFQMKFFYISQ 172
Db 127 ESPH-QMFTVYISIAHAFYIVSER-LEDFSEKSWLGYPTHEHRVMSAAAYKLYFIQI 184
QY 173 AYNFHSPELYFQVRKQDIPGQLIYIGLHFGGAYLLYLNHLGLLMLHYAVELL 232

Db 185 SYWIRHQFPEFYLOKLRKDRBIROKSQVAILHIAFISIAFYFNFTRVGLALITLEYITQLIF 244
QY 233 SVCSLLYFGDERYQKGLS-----LWPIVFIISGRVYTLIVSVYT--VGLHLACTNRNGNAL 285
Db 245 HIRAFHVG---RGLSDPAKFLNGSVLVRGLSGIIIVMTFWYGLRQA-ESFPVDIS 300
QY 286 SGNVNVLAAKIAVLSSGSIQYIITWTITVVLQWRLELDANLHVGCRKRRSRSGKTENG 345
Db 301 AGNFTAVIRNLVALLVQLFLYSFVW-----FHM-GRFPESNAKREKKKS 348
QY 346 VENPNRIDSPPKKEK 361
Db 349 AAAAAAV---PKKERK 361

RESULT 5
S46800
LAG1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YHL003c
C:Species: Saccharomyces cerevisiae
C:Date: 28-Oct-1994 #sequence_revision 28-Oct-1994 #text_change 23-Mar-2001
C:Accession: S46800; A54012
R:Favetto, T.
submitted to the EMBL Data Library, June 1994
A:Description: The sequence of S. cerevisiae cosmid 9780.
A:Reference number: S46797
A:Accession: S46800
A:Molecule type: DNA
A:Residues: 1-411 <FAV>
A:Cross-references: EMBL:U10555; NID:q500813; PIDN:AAB684429.1; PID:q500820; MIPS:YHL003c
R:D'mello, N.P.; Childress, A.M.; Franklin, D.S.; Kale, S.P.; Pinswasdi, C.; Jazwinski, J. Biol. Chem. 269, 15451-15459, 1994
A:Title: Cloning and characterization of LAG1, a longevity-assurance gene in yeast.
A:Reference number: A54012; MUID:94253121
A:Accession: A54012
A:Molecule type: DNA
A:Residues: 1-172, 'IV', 175-219, 'C', 221-300, 'TEISGI', 314, 'EKQE', 315, 'DSNDNPTE', 324, 'A', 324
'A', 381, 'AGOR', 386, 'L', 388, 'NRLARNEK' <DAM>
C:Cross-references: GB:U08133
C:Genetics:
A:Gene: SGD:LAG1
A:Cross-references: SGD:S0000995; MIPS:YHL003c
A:Map position: 8L
C:Function:
A:Description: involved in dertermination of longevity
C:Superfamily: hypothetical protein YKL008c
C:Keywords: transmembrane protein

Query Match 7.8%; Score 148.5; DB 2; Length 411;
Best Local Similarity 22.0%; Pred. No. 3.6e-05;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;
QY 26 MYSCVGMFFVLGLMFEPTAEMSIIVLTQLQGVVPAEGLPSGSRITLYHYGVKDLATVFY 85
Db 91 LVCVVSAYFLSGNRTESNPLHMFVAISQ-----VCGTDS-----YAKGKDLSEVFY 139
QY 86 MLVAIIIIATQYEVLDKLSRLRLTKGQKNLEAGQSLSVFYI-VSGTWGMIIIL-ASEN 143
Db 140 MIFFTFLREFLMDVIRPPTVYLVNTSEHRQRMLEQMYAIFCVGSGPFGLYIMYHSDL 199
QY 144 CLSDPTLLWKSQPHNMFMQKFFYISQLAYW-----FHSFPE 181
Db 200 WLFTKPMYRTPVTNTNPLFKIFIFLQGAFAWAQACVLVQLKPRKDYKELVFHHIYT 259
QY 182 L-----YFQYRKQDIPGQIYI-----GLHLFPH 205
Db 260 LLLIWSYVFHETKM-----GLAIYITMDVSDFFLSLTKTLNLSVPTPFVGLFVP- 312
QY 206 IGGAYLLYNHLGLLLMLHYAVELLSSVSLLYFGDERYQKGLSLWPIVFIISGRVLTLI 265
Db 313 ----FWIVLRHVNVIRILWSVLTEFRHEGNYVNLNATQYKWCISL-PIVFV-----LI 361

QY 266 VSVVTVGLH 274
Db 362 AALQLVNLY 370
RESULT 6
S30134
hypothetical protein YKL008c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL156
C:Species: Saccharomyces cerevisiae
C:Date: 18-Jun-1993 #sequence_revision 18-Jun-1993 #text_change 24-Sep-1999
C:Accession: S30134; S37819; S17017
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Dujon, B.
Yeast 9, 279-287, 1993
A:Title: Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI reveals
mammalian UOG-1 gene.
A:Reference number: S30132; MUID:93255906
A:Accession: S30134
A:Molecule type: DNA
A:Residues: 1-418 <BOY>
A:Cross-references: GB:S59773; NID:g300231; PIDN:AAC60549.1; PID:g300232
A:Experimental source: strain S288C
R:Boyer, J.; Pascolo, S.; Richard, G.F.; Ghazvini, M.; Collea, L.; Thierry, A.; Men
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37813
A:Accession: S37819
A:Molecule type: DNA
A:Residues: 1-418 <BO2>
A:Cross-references: EMBL:Z28008; NID:g485982; PIDN:CAA81843.1; PID:g485983; MIPS:YKL0
A:Experimental source: strain S288C
R:Amatruda, J.F.; Gattermeier, D.G.; Cooper, J.A.
submitted to the EMBL Data Library, August 1991
A:Description: Yeast capping protein.
A:Reference number: S17016
A:Accession: S17017
A:Molecule type: DNA
A:Residues: 1-149 <AMA>
A:Cross-references: EMBL:X61398; NID:g455515; PIDN:CAA43670.1; PID:g3445
C:Genetics:
A:Map position: 11L
C:Superfamily: hypothetical protein YKL008c
C:Keywords: transmembrane protein

Query Match 7.2%; Score 137; DB 2; Length 418;
Best Local Similarity 23.5%; Pred. No. 0.00037;
Matches 61; Conservative 44; Mismatches 93; Indels 62; Gaps 11;
QY 67 GSRTLYHYGVKDLATVFYVYMLVAIIIIHATIQEYVLDKLSRRLQLTGKQKNLEAGQSLV 126
Db 121 GDTNAYKGINDLCFVYFYIMIFFTLREFLMDVIRPFAIRLHVTSKHKRKRIMEQYAI 180
QY 127 FYI-VSGIWMIIIL-ASENCLSDPTLLWKSQPHNMFMQKFFYISQLAYW----- 175
Db 181 FYTGVSGPGFICMIVHSDLWFFNTKAMYRTPDTPNPELFKVFYLGAAFAWAQACILVL 240
QY 176 -----FHSFPEL-----YFQYRKQDIPGQIYIIGLHFIHGAYLIYNH 216
Db 241 QLEKPRKDHNEITFHIIIVTLLIWSYVFHETKMGLP---IYITMDVSDFFLSFSTLNY 297
QY 217 L--GLL-----LLMLHYA-VELLSVCS-----LLYFGDERYQKGLSLWPI 254
Db 298 LDSGLAFFSFAIVVAMIYLRHYINLILWSVLTPQRTGNYVNLNATQYKWCISL-PI 356
QY 255 VFISGRVTLIVSVTVGLH 274
Db 357 VFV-----LIGALQLVNLY 370

RESULT 7
T40389
longevity assurance protein homolog SPBC3E7.15c - fission yeast (Schizosaccharomyces
C:Species: Schizosaccharomyces pombe

C; Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C; Accession: T40389; T40499
R; lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Brown, D.; Churcher, C.M.
submitted to the EMBL Data Library, May 1998
A; Reference number: Z21924
A; Accession: T40389
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-357 <LYN>
A; Cross-references: EMBL:AL023534; PIDN:CAA19018.1; GSPDB:GN00067; SPDB:SPBC3E7.15C
A; Experimental source: strain 972h; cosmid c3E7
R; Gwilliam, R.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A; Reference number: Z21933
A; Accession: T40499
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-357 <GWI>
A; Cross-references: EMBL:AL011534; PIDN:CAA20722.1; GSPDB:GN00067; SPDB:SPBC4F6.02C
A; Experimental source: strain 972h; cosmid c4F6
C; Genetics:
A; Gene: SPBC3E7.15C; SPBC4F6.02C
A; Map position: 2

Query Match 7.0%; Score 133; DB 2; Length 357;
Best Local Similarity 20.5%; Pred. No. 0.00068;
Matches 48; Conservative 45; Mismatches 117; Indels 24; Gaps 7;

QY	41	EGTAEISIVLTL-----QHGVVPAEGLPGSRTLYHYGVKDLATVFYFYL	87
		: : :	
Db	29	EKTWIVPLIILLVLGVFYNPNGYIKYIGFL-SYPIGTNPQYQKGRDIAFCFLYAL	87
		: : :	
QY	88	VAILIHATICEYVLDKLSRLQL-TKGQKNLNEAGOLSVYIVSGIWGMIL-ASENCL	145
		: : :	
Db	88	FFTCREFIMEIARIGRHFNINAPAKLRFEQAYTCLIFVTGWSGLYVNMKQPMWF	147
		: : :	
QY	146	SDPTLLWKSOPHNMTFMQKFFYISOLAYFHSFPPELYFO--KVRQDIPGOLYIGLHL	203
Db	148	FNTDAFWEYEPHYVHVGSRFAYLIEAAWYIQALVLLQLKPRK-DPELIVVHHITL	206
		: : :	
QY	204	FHTGGAYLLYNIHGLLLMLHYAVELLSSVCSLLYFGDERYQKGLSLMPIVEI	257
		: : :	
Db	207	LLGLSYFHTFTGLAVFTMDTSDIWLALSCKL-----NYNTVTVPIPIFYI	255

RESULT 8
T38012
longevity-assurance protein 1 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38012
R;Churche, C.M.
submitted to the EMBL Data Library, September 1997

A;Accession: J36012
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-390 <CHU>
A;Cross-references: EMBL:299258; PIDN:CAB16359.1; GSPDB:GNO00066; SPDB:SPAC1A6.09c
A;Experimental source: strain 972h⁻; cosmid clA6
C;Genetics:
A;Gene: lagI; SPDB:SPAC1A6.09c
A;Map position: 1

Query Match 6.6%; Score 125; DB 2; Length 390;
Best Local Similarity 23.9%; Pred. No. 0.0037;
Matches 57; Conservative 35; Mismatches 100; Indels 46; Gaps 11;

[illegible]

Qy	120	EAGQLSVFYIVSGIHWGMILLIASENCLSDPTLLWKSQPHNMNTQMEFFYSILAYWFHSF	179
Db	157	EQGVSFFYICFVFWLGLYIYTRSSNWSNEEKLFEDIPQYYMSPFLFRAYLILQGFNLQOI	216
Qy	180	PELYQKVRKOD-----IPQLIYI--GLHLPFHIGGAYLLYNHIGLGLL-----ML	224
Db	217	LVLHLEQ--RRADHWQMFAMHIVTCAILLISYGFNFRVGNQ--ILYFDLSDYILSGKML	274
Qy	225	HYAVELLSVCSLLIYFGDERYQKGLSLWPIVFISGRLLVTLIVSVWV-----GHLH	275
Db	275	KYLQ--EGKICDYL-----GIFVASWVYSRHYLFESKILRVVYTNAPETIIGGPHL	322

RESULTS

B86726
hypothetical protein yicA [imported] - Lactococcus lactis subsp. lactis (strain IL149)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: B86726
P:Belotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eil
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471

Query Match	6.2%;	Score 116.5;	DB 2;	Length 696;
Best Local Similarity	21.9%;	Pred. No. 0.041;		
Matches 80;	Conservative 62;	Mismatches 122;	Indels 101;	Gaps 18;

QY 86 MLVAIIIHATIQEYVLDKLSRRLQLTKGKQKLNKAGOLSVFYIVSGIWGMIIASENCL 145
 || | : | : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
 Db 75 ML----IWNARMTSIIFEIFA--MOIRKGLFNIN-----SLIYVILGLLINVLVSGKKAF 124

OV 146 SPPTLLWKSOPHNMMTFOMKFFYISOLA-----YWFHSFPELYFKVRKQDIPG 194

—
—
..

—
—
..

—
—
..

—
—
..

..
..
..
—
..
..
—
—
..
—

..
—

Db: 125 LKPS-----HLSLTFLLMWFFLP
125 LKPS-----HLSLTFLLMWVSGATN

09 195 01.TVTG1H1E-----HTGG--AVT.I.VT.-----NH-----I.GI.I.I.I.M.H 225

Variable	Mean	SD	Min	Max
Age	35.5	10.5	20	55
Gender	0.5	0.5	0	1
Marital Status	0.5	0.5	0	1
Education	12.5	1.5	10	15
Income	3000	1000	1000	5000
Health	0.5	0.5	0	1
Stress	0.5	0.5	0	1
Life Satisfaction	0.5	0.5	0	1

Db 178 RSNWISLGLFILGLTGLTNEVGATAFLALLETIFYNRRQPSERVLTQIFGVLGAGIG 237

017 226 YAVET I SSV-----CSL I YECDEPYOKC I SIWPIVETSCRIYTI I VS- 267

[illegible]

Db 238 FFIQLLSSGSSETQNYGKSAGFLQHLSDVFTGTMQY-SGFLLLPILLGGLYLRRIQW 296

Q7 368 -----VATWVCY WT ACENPNCNAI SCNBNNI AAVT - - AVI SSSCSCTOVVITWTETI TTVVLT. 318

```
QY Z08 -----VIVGLALAGINKNNGNALSNGNVNVLAKRT AVSSSCSIQVIITIIWIDIIIVME JTO
```

Db 297 TEKVKTIVTSLFL-----GSALAGSIAILASPISPARLWFAPNILLIITLLLL--I 347

Q: 310 CONT P 333

QY 319 QRWLE 323
: | |

Db 348 EAWQE 352

RESULT 10

A69845.

Na⁺/H⁺ antiporter homolog yjbQ - *Bacillus subtilis*

C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence revision 05-Dec-1997 #text change 15-Oct-1999

C;Accession: A69845

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

Genetics:

Search completed: September 6, 2002, 17:11:51
Job time: 6790 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:10:42 ; Search time 26.42 Seconds
(without alignments)
531.991 Million cell updates/sec

Title: US-09-807-470-2

Perfect score: 1892

Sequence: 1 MGLRKKRNAPNVLPSHEFWV.....NGVENPNRIDSPPKKKEKAP 363

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1145.5	60.5	373	1	TRAM_CANFA
2	1129.5	59.7	373	1	TRAM_HUMAN
3	1084.5	57.3	358	1	TRAM_BOVIN
4	754	39.9	370	1	Y557_HUMAN
5	148.5	7.8	411	1	LAG1_YEAST
6	137	7.2	418	1	YKAB_YEAST
7	133	7.0	384	1	YHXP_SCHPO
8	125	6.6	390	1	LAG1_SCHPO
9	97.5	5.2	533	1	GLPT_HUMAN
10	97	5.1	1450	1	CFTR_RABIT
11	96.5	5.1	534	1	COX1_RICPR
12	95	5.0	370	1	YDIK_ECOLI
13	94.5	5.0	540	1	SOAL_MOUSE
14	94.5	5.0	788	1	QOXM_SULAC
15	94	5.0	272	1	ATP6_BUGAP
16	93.5	4.9	521	1	COX1_APILI
17	93	4.9	532	1	YHCA_BACSU
18	91.5	4.8	369	1	Y316_MYCPN
19	91	4.8	307	1	OCDD2_HUMAN
20	91	4.8	624	1	NKX3_RAT
21	91	4.8	645	1	NKX3_MOUSE
22	90	4.8	513	1	COX1_RABIT
23	90	4.8	514	1	COX1_CERSI
24	90	4.8	514	1	COX1_EQUAS
25	90	4.8	514	1	COX1_HORSE
26	90	4.8	514	1	COX1_RHUN
27	90	4.8	516	1	COX1_GADMO
28	89	4.7	360	1	MRAY_PASMU
29	89	4.7	430	1	SECY_BACHD
30	89	4.7	499	1	MYIN_AQUAE
31	89	4.7	514	1	COX1_BOVIN
32	89	4.7	514	1	COX1_CANFA
33	89	4.7	514	1	COX1_FELCA

34	89	4.7	514	1	COX1_SHEEP	078749 ovis aries
35	89	4.7	644	1	NKX3_HUMAN	09hc58 homo sapien
36	89	4.7	649	1	CNTL_HUMAN	000337 homo sapien
37	88.5	4.7	1481	1	CFTR_BOVIN	P33071 bos taurus
38	88.5	4.7	1481	1	CFTR_SHEEP	000555 ovis aries
39	88	4.7	433	1	SECY_RICPR	Q9zcs5 rickettsia
40	88	4.7	515	1	COX1_HIPAM	Q9zzy9 hippopotamu
41	88	4.7	531	1	YDFG_SCHPO	010487 schizosacch
42	87.5	4.6	297	1	CTAA_BACFI	004443 bacillus fi
43	87.5	4.6	515	1	COX1_CHICK	P18943 gallus gall
44	87	4.6	432	1	REBX_SALTI	Q91911 salmoneilla
45	87	4.6	514	1	COX1_PIG	079876 sus scrofa

ALIGNMENTS

```

RESULT 1
ID      TRAM_CANFA
AC      Q01685;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DE      16-OCT-2001 (Rel. 40, Last annotation update)
DE      TRAM protein (Translocating chain-associating membrane protein).
OS      Canis familiaris (Dog).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX      NCBI_TaxID=9615;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 1-26 AND 164-184.
RC      TISSUE=Kidney;
RX      MEDLINE=92244357; PubMed=1315422;
RA      Goerlich B., Hartmann E., Frenn S., Rapoport T.A.;
RT      "A protein of the endoplasmic reticulum involved early in polypeptide
      translocation.";
RL      Nature 357:47-52(1992).
CC      -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF
      SECRETORY PROTEINS ACROSS THE ER MEMBRANE.
CC      -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.
CC      -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
      the European Bioinformatics Institute. There are no restrictions on its
      use by non-profit institutions as long as its content is in no way
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
      or send an email to license@isb-sib.ch).
      -----
      EMBL: X63678; CAA45217.1; -.
      PIR: S21736; S21736.
      Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
      INIT_MET 0
      DOMAIN 1 28
      TRANSMEM 29 49
      DOMAIN 50 75
      TRANSMEM 76 96
      DOMAIN 97 120
      TRANSMEM 121 141
      TRANSMEM 142 158
      TRANSMEM 159 179
      TRANSMEM 180 191
      TRANSMEM 192 212
      TRANSMEM 213 216
      TRANSMEM 217 237
      TRANSMEM 238 250
      TRANSMEM 251 271
      TRANSMEM 272 296
      TRANSMEM 297 317
      TRANSMEM 318 373
      CARBOHYD 55
      SEQUENCE 373 AA; 43029 MW; 1D85808E1D80E835 CRC64;

```



```
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark T.G., Morris J., Akamatsu M., McGraw R.A., Ivarie R.D.;
RT "Cloning and sequence analysis of a bovine tram cDNA.";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLLOCATION OF
CC SECRETORY PROTEINS ACROSS THE ER MEMBRANE (By similarity).
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19578; AAG10391.1; -
KW Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
FT NON_TER 1
FT DOMAIN <1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 60 LUMENAL (POTENTIAL).
FT TRANSMEM 61 81 POTENTIAL.
FT DOMAIN 82 105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 106 126 POTENTIAL.
FT DOMAIN 127 143 LUMENAL (POTENTIAL).
FT TRANSMEM 144 164 POTENTIAL.
FT DOMAIN 165 176 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 177 197 POTENTIAL.
FT DOMAIN 198 201 LUMENAL (POTENTIAL).
FT TRANSMEM 202 222 POTENTIAL.
FT DOMAIN 223 235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 236 256 POTENTIAL.
FT DOMAIN 257 281 LUMENAL (POTENTIAL).
FT TRANSMEM 282 302 POTENTIAL.
FT DOMAIN 303 358 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC... ) (PROBABLE).
SQ SEQUENCE 358 AA; 41403 MW; 25CF9930C4CDDA15 CRC64;

Query Match
Best Local Similarity 57.3%; Score 1084.5; DB 1; Length 358;
Matches 214; Conservative 53; Mismatches 79; Indels 11; Gaps 5;

QY 17 EFVQNHADVMSCVGMFFVLGLMFEFTAEMSIVFLTLQHGCVVPA-EGLPSSGRTLYHYG 75
DB 1 EFVLQNHADIVSCVAMVFLGLMFEITAKVSIIFVLQYVNTLPATYEQNTESAFLYYIG 60
QY 76 VKDLATVFFVYMLVAIIHATIQEYVLDKLSRRQLTKGKQKLNKAGOLSVFVIVSGIWG 135
DB 61 IKDLATVFFVYMLVAIIHATIQEYVLDKLNRRMHFSKTKHKNESQLSAFYLFSCIWG 120
QY 136 MIILASENCSDPTLLWKSPHNMFTQMKFFYISQLATWFHSPFELYQKVRQDIPGQ 195
DB 121 TFLISENIDPTLLWRAYPHNLMTFQMKFFYISQLATWFHAFPELYFQTKTKEDIPRQ 180
QY 196 LIYIGLHLPHIGAYLLYNHLGLLLMLHYAVELSSVCSLYFGDERYOKGLSLWPV 255
DB 181 LVYIGLYLPHIAGAYLLNLNLHGLVLLVLYHVFELHISRLYFFDEKQKGFSLWAVL 240
QY 256 FTSGRLVTLTVSVTVGLHLA-GTNRGNALSGNVNLAALAVLSSSCSIQYITWTTL 314
DB 241 FVLGRLLTLTVLTVGLGARAENQKLPSTGNVNLAVRIAVLASICITQAFMMWKEI 300
QY 315 TWVLQRLWLEDANLHVCGRRK-----RSRS-RKGTENGVE----NPNRIDSPPKKKEA 362
DB 301 NFQRLRWREHSAPQAPAVKKKPPVTKGRSXXKGTENGVTGTSNGADSPNRNKRKS 357
```

```
RESULT 4
Y557_HUMAN STANDARD; PRT; 370 AA.
AC Q15035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein KIAA0057.
GN KIAA0057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow; PubMed=7584044;
RX MEDLINE=96051398; Miyajima N., Suzuki T., Tanaka A., Sato S.,
RA Nomura N., Nagase T., Ishikawa K.-I., Tabata S.,
RA Seki N., Kawarabayashi Y.,
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RN Tracey A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31762; BAA06540.1; -
DR EMBL; AL049611; CAB71119.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
SQ SEQUENCE 370 AA; 43327 MW; 9B5183F1A3D45366 CRC64;

Query Match
Best Local Similarity 39.9%; Score 754; DB 1; Length 370;
Matches 158; Conservative 42.4%; Pred. No. 1.1e-55;
Matches 158; Conservative 63; Mismatches 132; Indels 20; Gaps 5;

QY 5 KKNARPPVLSHFQVONHADVMSCVGMFFVLGLMFEFTAEMSIVFLTLQHGCVVPAEGL 64
DB 4 RRTKSYPLFSQEFVHNHADIGFCLVLCVLGLMFEFTAEMFTAPFLFLPQINISVPT- 60
QY 65 PSCSRTL-YHYGVKDLATVFFVYMLVAIIHATIQEYVLDKLSRRQLTKGKQKLNKAGQ 123
DB 61 -ADSETVHYHYGPKDLVTILFYITILHVVQVEYILDKISRLHLSKVHKSFNESGQ 119
QY 124 LSVFYIVSGTWGMIILASENCSDPTLLWKSPHNMFTQMKFFYISQLATWFHSPFELY 183
DB 120 LVVFHFTSVIWCFFVYVVTGTYLTPRSLWEDYDPVHLPFQVKKFFYLCQLAYLHALPELY 179
QY 184 FQVRQDIPQILYICLHLPHIGAYLLYNHLGLLLMLHYAVELSSVCSLYFGDE 243
DB 180 FQVRKEEIPRQYICLYLVHAGAYLLNLNLHGLVLLVLYHVFELHISRLYFFDEKQKGFSLWAVL 239
QY 244 RYQGLSLWPVIVTSGRLVTLVSVTVGLHLA-GTNRGNALSGNVNLAALAVLSSS 302
```

[illegible]

RESULT	5
LAGI_YEAST	
ID	LAGI_YEAST STANDARD; PRT; 411 AA.
AC	P38703;
DT	01-FEB-1995 (Rel. 31, Created)
DT	11-FEB-1995 (Rel. 31, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Longevity-assurance protein 1 (Longevity assurance factor 1).
GN	LAG1 OR YHJ003C.
OS	Saccharomyces cerevisiae (Baker's yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX	NCBI_TaxID=4932;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=X2180-1A;
RX	MEDLINE=94253121; PubMed=8195187;
RA	D'Mello N.P., Childress A.M., Franklin D.S., Kale S.P.,
RA	Pinswasi C., Jazwinski S.M.;
RT	"Cloning and characterization of LAG1, a longevity-assurance gene in
RT	yeast.;"
RL	J. Biol. Chem. 269:15451-15459(1994).

RC	STRAIN=S288C / AB972;
RX	MEDLINE=94378003; PubMed=8091229;
RA	Johnston M., Andrews S., Brinkman R., Ding H., Dover J.,
RA	Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA	Kucaba T., Hiller L., Jier M., Johnston L., Langston Y.,
RA	Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.
RA	Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA	Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA	Vaudin M.;
RT	"Complete nucleotide sequence of <i>Saccharomyces cerevisiae</i> chromosome
RT	VIII.";
RL	Science 265:2077-2082(1994).
CC	- !- FUNCTION: INVOLVED IN THE AGING PROCESS. DELETION OF LAG1 RESULTS
CC	IN A PRONOUNCED INCREASE (APPROXIMATELY 50%) IN MEAN AND IN
CC	MAXIMUM LIFE SPAN.
CC	- !- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC	- !- SIMILARITY: BELONGS TO THE LAGS1 FAMILY

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC	Sequence	POTENTIAL
EMBL	U081133; AAA21579.1; -	POTENTIAL.
DR	EMBL; U10555; AAB68429.1; -	POTENTIAL.
DR	PIR; S46800; S46800.	POTENTIAL.
DR	SGD; S000095; LAG1.	POTENTIAL.
KW	Transmembrane.	POTENTIAL.
FT	TRANSMEM 82	POTENTIAL.
FT	TRANSMEM 135	POTENTIAL.
FT	TRANSMEM 177	POTENTIAL.
FT	TRANSMEM 212	POTENTIAL.
FT	TRANSMEM 252	POTENTIAL.
FT	TRANSMEM 297	POTENTIAL.
FT	TRANSMEM 356	POTENTIAL.

FT	CONFLICT	173	174	ML-> IV (IN REF. 2).
FT	CONFLICT	220	240	F -> C (IN REF. 2).
FT	CONFLICT	301	411	
FT				VFDPFGLVFFVFFVHVHVNVNIRILMSVLTPEFHGNYVL
FT				NFAQYQKCTSLSPFVFLAALQNLVLYFLILRLYL
FT				IWQIQKDRSDSDSEANESEKCE -> TEISGIWE
FT				KQEDNSNPTRALSPNETSKQVDPDLLVLPNTENRAL
FT				LEAKSRPTTITATDSEPSLVTPYIPGNDSDLSRNVNLL
FT				GLVARAQGLQNLARNEK (IN REF. 1).
SQ	SEQUENCE	411	AA: 48454	91676D56AC053F3C CRC64:
			GW:	

```

Query Match      7.8%; Score 148.5; DB 1; Length 411;
Best Local Similarity 22.0%; Pred. No. 3.7e-05;
Matches 68; Conservative 46; Mismatches 106; Indels 89; Gaps 11;

QY 26 MVSQVGMFFVLGLMFECTAEMSIYFLRLQHCVVVPAECLPSGSRTHLYHYGVKDLATVFFY 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 91 LVCVSYAFUSGNTESNPJLHMFAISYQ-----VDGTD-----YAKGKIDLSVFFY 139
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 86 MLVAIIITHATIQEVLVDKLSRRQLQTKGKNKEAGQLSVFYI-VSGIMGMIIL-ASEN 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 MIFFTLREFLMDVVRPFVYLVNVTSEHQKRMLEQMYAIFYCVSGPGLIYWIHSDL 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 144 CLSDPTLLWKSQPHNMWTFQMKFFYISOLAYW-----FHSFPE 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 WLFTKPMYRTYPITNPFLEKFIFYLQAAFAWQAQACVLVLEKPRKDYKELVFHHVLT 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 182 L-----YFQVKRQDIPGQLIYI-----GLHLFH 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 LLLIWSSYVFHTKM-----GLAIYITMDVSDFFLSKTLNLYNSVFTPFVGLFV 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 206 IGGAYLLYLNLHGLLLMLHYAVELLSSVCSLFFYGBERYQGLSLMIPFIYISGRVTLI 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 ----FWIYLHRVYNIRILMSVLFREHGNVYLVNFATQYKWCISL-PIYFV-----LI 361
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 266 VSVVTVGLH 274
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 AALQVLNLY 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      6
YKAB_YEAST
ID          YKAB_YEAST      STANDARD;      PRT;      418 AA.
AC          P28496;
DT          01-DEC-1992 (Rel. 24, Created)
DT          01-JUN-1994 (Rel. 29, Last sequence update)
DT          16-OCT-2001 (Rel. 40, Last annotation update)
DE          Hypothetical 49.0 kDa protein in UFD4-CAP1 intergenic region.
GN          YKL008C OR YKL156.
OS          Saccharomycetes cerevisiae (Baker's yeast).
OC          Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC          Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TAXID=4932;
[1]
CBLI
RN          RN          SEQUENCE FROM N.A.
RP          MEDLINE=93255906; PubMed=8488728;
RA          Boyer J., Pascolo S., Richard G.F., Dujon B.;
RA          "Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI
RT          reveals four open reading frames, including the CAP1 gene, an intron-
RT          containing gene and a gene encoding a homolog to the mammalian UOG-1
RT          gene.";
RT          Yeast 9:279-287(1993).
RL          [2]
RN          SEQUENCE OF 1-149 FROM N.A.
RP          MEDLINE=93077675; PubMed=1447293;
RA          Amatruda J.F., Gattermeir D.J., Karpova T.S., Cooper J.A.;
RA          "Effects of null mutations and overexpression of capping protein on
RT          morphogenesis, actin distribution and polarized secretion in yeast.";
RT          J. Cell Biol. 119:1151-1162(1992).
RL          CC          -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC          -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC          -----
CC          This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X61398; CAA43670.1; -;
DR EMBL; S59773; AAC60549.1; -;
DR EMBL; Z28008; CAA81843.1; -;
DR PIR; S30134; S30134;
DR SGD; S0001491; YKL008C;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
SQ SEQUENCE 418 AA; 48992 MW; 7691BAG23AC0460A CRC64;

Query Match 7.2%; Score 137; DB 1; Length 418;
Best Local Similarity 23.5%; Pred. No. 0.00034;
Matches 61; Conservative 44; Mismatches 93; Indels 62; Gaps 11;

QY 67 GSRTLYHYGVKDLATVFFYVMAIIHATIQEYVLDKLSRRLOLTGKQKNEAGLSV 126
DB 121 GDTNAYGKGLNDLCFFYYIFFFLREFLMDVVRPFAIRLHVTSKHKRIKMEQWAI 180
QY 127 FYI-VSGINGMIL-ASENCLSDPTLLWKSOPHNMTFMKFFYISOLAYV----- 175
DB 181 FYTGVSQPGFYCMYHSDLWFNTKAMRYTPDFTNPFKVFYLGQAFAWAQACILVL 240
QY 176 -----FHSPFEL-----YFQVKRKODIPGQLIYIGLHLFHGGAYLLYNH 216
DB 241 QLEKPRKHNLTFHHIVTLLLSWSSYVFHTKMGLP---IYITMDVDFLLSFKTLNY 297
QY 217 L--GLL-----LLMLHYA-VELLSVCS-----LLYFGDERYQKGLSLWPI 254
DB 298 LDSGLAFFFAIFVAVIYLRHVINLKILWSVLTFQFTGNYVNFATQYKWCWISL-PI 356
QY 255 VFISGRRLTVSVTVGLH 274
DB 357 VFV-----LIGALQLVNLV 370

RESULT 7
ID YHXF_SCHPO STANDARD; PRT; 384 AA.
AC O59735;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 42.4 kDa protein C3E17.15c in chromosome II.
GN SPBC3E7.15C OR SPBC4F6.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Brown D.,
RA Churcher C.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Gwilliam R., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE LASSI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL023534; CAA19018.2; -;
DR EMBL; AL031534; CAA20722.2; -;
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 233 253 POTENTIAL.
FT TRANSMEM 270 290 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
SQ SEQUENCE 384 AA; 45335 MW; 82FCF8EA6638849A CRC64;

Query Match 7.0%; Score 133; DB 1; Length 384;
Best Local Similarity 20.5%; Pred. No. 0.00067;
Matches 48; Conservative 45; Mismatches 117; Indels 24; Gaps 7;

QY 41 EGTAEMSIIVLT-----OHGVVVPAGEPLSGSRTLYHYGVKDLATVFFVML 87
DB 56 EKTWVPLILLTLVGWYFVNPNGYIKYIGFL-SYIPGTNPAGYKGRDLIAFLFYAL 114
QY 88 VAIITHATIQEYVLDKLSRRLOLTGKQKNEAGLSVFIYVSGIWGMIL-ASENCL 145
DB 115 FFFCFREFIMQEIIRIGRHNIRAPAKLRFEQATCLYFTYVMSGLYVMKQTPWF 174
QY 146 SDPTLLWKSOPHNMTFMKFFYISOLAYWPHSPPELYFQ--KVRKODIPGQLIYIGLHL 203
DB 175 FNTDAFWEYEPHFYHVGVSFAFYIEAAYWTOQALVILQLEKPRK-DFRELVVHHITL 233
QY 204 FHGGAYLLYNHGLLLMLHYAVELLSSVCSLLYFGDERYQKGLSLWPIVFI 257
DB 234 LLIGLSYFHTWIGLAVFITMDTSDIWLALSKCL-----NYVNTVIVIPFI 282

RESULT 8
ID LAGL_SCHPO STANDARD; PRT; 390 AA.
AC P78970; O13860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR SPAC1A6.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Chanda E.R., Lingner C., Ko Z., Young P.G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN THE AGING PROCESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE LASSI FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

Db 102 SGDWTLPDLA-----WLNTP-----VIGAKLYAGV 127
QY 202 H-LFHIGGAYLL-----YLN-----HLGLLLMLHYAVELSSVCSLLYFGDER 244
Db 128 HNLDMGCTAIMAKVRPIYTTTTTWFVGAHQHIG--RFMHVCLMLLFS--ALLYMRGEQ 183
QY 245 YQKG-----LSLWPIFIVS-GRVTLIVSVVTVGLHLAGLNRNGNAL 285
Db 184 VAQCIIRHFATRLAGVRGDAVLLAAQAIRAVAGLVVVVTVLVAQVLGGI-----GLAV 235
QY 286 SGNNVLAAKIIVLSSCSIQ-----VYIWTWLTIVV 317
Db 236 SG-VPYATLLTVLMILSCLVOLGPLPVLPAIIWLYWTGDTTW 277

RESULT 13
SOAL_MOUSE
ID SOAL_MOUSE STANDARD; PRT; 540 AA.
AC Q61263; Q64180;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase
1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
GN SOAT1 OR ACAT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=96064687; PubMed=7592824;
RA Uelman P.J., Oka K., Sullivan M.C., Chang T.-Y., Chang C.C.Y.,
Chan L.;
RT Tissue-specific expression and cholesterol regulation of acylcoenzyme
A:cholesterol acyltransferase (ACAT) in mice. Molecular cloning of
mouse ACAT cDNA, chromosomal localization, and regulation of ACAT in
vivo and in vitro.;
RL J. Biol. Chem. 270:26192-26201(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96158986; PubMed=8579615;
RA Green S., Steinberg D., Quehenberger O.;
RT "Cloning and expression in Xenopus oocytes of a mouse homologue of the
human acylcoenzyme A: cholesterol acyltransferase and its potential
role in metabolism of oxidized LDL";
RL Biochem. Biophys. Res. Commun. 218:924-929(1996).
CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL
ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
CHOLESTEROL ABSORPTION.
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
ester.
CC -!- SUBCELLULAR LOCATION: Integral membrané protein. Endoplasmic
reticulum.
CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; L42293; AAC42075.1; -;
DR EMBL; S81092; AAB36050.1; -;
DR MGD; MGI:104665; Soat1.
DR InterPro; IPR002688; ACAT.
DR Pfam; PF01800; ACAT; 1.
KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
KW Cholesterol metabolism.
FT TRANSMEM 132 152
POTENTIAL.

FT TRANSMEM 311 331
FT TRANSMEM 352 372
FT TRANSMEM 460 480
FT TRANSMEM 488 508
FT TRANSMEM 495 515
SQ SEQUENCE 540 AA; 63739 MW; 8EF900C8BCDF73C0 CRC64;
Query Match 5.0%; Score 94.5; DB 1; Length 540;
Best Local Similarity 21.5%; Pred No 1.6;
Matches 45; Conservative 34; Mismatches 93; Indels 37; Gaps 6;
QY 48 IVFELTQHGVS-----VPAEGLPSGSRSLTYHYGVKDLATVFFYMLVAIIHATIQEYV 100
Db 215 LLFLVFGVGLGVFVPTVWVLAAYLTPPASR-----FILILEQIRLIMKAHSFV 261
QY 101 LDKLSRLQLTKGQNK--LNEAGOLSVFYIVSGINGMILASENCISLDPILLKWSOPHN 158
Db 262 RENIPRVLNAAKESSKSDPLFTVNOYLYF-----LFAPTLYRDNYPRTPVVRW---GYV 313
QY 159 MMTFOMKFFVYSOLAYWFHSFPPELYFOKVRKODIPGOLYIGLHFGHGGAYLLYLNLHLG 218
Db 314 AMQFLQVFGCLFYVYVIFERLCAPLFRNIQKQEPFSARVLVLCVENSILPGVLIILFLSFFA 373
QY 219 LLLMLHYAVVELSSVCSLLYFGDERYOK 247
Db 374 FLHCLWNAFAEMLR-----FGDRMFYK 395
RESULT 14
QOXM_SULAC
ID QOXM_SULAC STANDARD; PRT; 788 AA.
AC P39481;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Quinol oxidase polypeptide I/III (EC 1.9.3.-).
GN SOXM.
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX MEDLINE=94357214; PubMed=8076636;
RA Luebben M., Arnaud S., Castresana J., Warne A., Albracht S.P.J.,
Saraste M.;
RT "A second terminal oxidase in Sulfolobus acidocaldarius";
RL Eur. J. Biochem. 224:151-159(1994).
CC -!- FUNCTION: TERMINAL OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SOXM
FORMS THE FUNCTIONAL CORE OF THE ENZYME COMPLEX.
CC -!- FUNCTION: CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING
IN A QUINOL ARE TRANSFERRED TO THE BIMETALLIC CENTER OF SOXM
FORMED BY A HEME AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferrocyclochrome
c + 2 H(2)O.
CC -!- COFACTOR: TWO HEME GROUPS AND COPPER B.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBUNIT: FORMS A COMPLEX WITH AT LEAST SOXM AND A 30 kDa RIESKE
FE-S PROTEIN, BUT NEITHER WITH SOXA NOR SOXB.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: SOXM IS PROBABLY A PRECURSOR FORM OF SUBUNIT I AND III.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE HEME-COPPER
RESPIRATORY OXIDASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE CYTOCHROME C
OXIDASE SUBUNIT 3 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC EMBL; X73567; CAA51969.1; -.
CC HSSP; P00396; LOCC.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR Pfam; PF00510; COX3; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
DR PROSITE; PS0253; COX3; 1.
KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain.
FT TRANSMEM 19 39
FT TRANSMEM 58 78
FT TRANSMEM 100 119
FT TRANSMEM 152 169
FT TRANSMEM 187 206
FT TRANSMEM 232 252
FT TRANSMEM 276 294
FT TRANSMEM 307 326
FT TRANSMEM 337 356
FT TRANSMEM 369 388
FT TRANSMEM 411 432
FT TRANSMEM 449 468
FT TRANSMEM 526 544
FT TRANSMEM 547 564
FT TRANSMEM 614 634
FT TRANSMEM 657 676
FT TRANSMEM 690 707
FT TRANSMEM 729 749
FT TRANSMEM 766 786
FT METAL 62 62
FT METAL 238 238
FT METAL 242 242
FT METAL 288 288
FT METAL 289 289
FT METAL 374 374
FT METAL 376 376
SQ SEQUENCE 788 AA; 87082 MW; 3939C16CDB8A08AD CRC64;

Query Match 5.0%; Score 94.5; DB 1; Length 788;
Best Local Similarity 19.5%; Pred. No. 2.5;
Matches 73; Conservative 57; Mismatches 98; Indels 147; Gaps 19;

QY 28 SCVG-MFVVLGL--MFGETAEMSIV-----FLTLQ-----HG-----VVVP-AE 62
DB 17 SDVGQMYIVLGIVALLIGSVNAALIRDSFNNLNAVVDYDAVTLHGIFMFFVYVPLST 76
QY 63 GLPS-----GSRTLYHYGVKDLATVFFYMLVAIIHATIQE-----YVLDKLSR 106
DB 77 GFANYLVPRMIGAHDLWPKINALS---FWMLVPAVILAALSPLLGAVDLGWYIAPLSV 133
QY 107 RLQLTGKQKNEAGQLSVFYIVSG-----IWGMII 138
DB 134 ETTVNVGLGTNL-----IQIALISGLSLTGTGVNFVMTIKMKVPYKMKPLFWGCFET 188
QY 139 LASENCLSDPILLWKSQPHNMFTQMKFFYSOL-----AYWHSF 179
DB 189 TAILMIAMPISL-----TAGLVFAYLERLWGTFFDSALGSGSPVLWQQLFWFGFH 238
QY 180 PELY-----FQKVRKODIPQQLIYIGLHPLFHIGGAYL---LYNLHLGLLL 222
DB 239 PEVYIILIPAMGLVSELPKARREIFG---YTAIALSSIAIALSALGVVWHH--WETA 293
QY 223 MLHAYVELLSVCSL-----LYFGDERYQKLSL---WPIVFISGRVLT 263
DB 294 IDNTLVQIVSSATTMAIAPSGVKVNLNTATLYGGEIRYKPTILLISFIVMFLGGITG 353
QY 264 LIVSVVTVGLHLACT 278
DB 354 VFFPLVPIDYALNGT 368
```

```
RESULT 15
ATP6_BUCAP STANDARD; PRT; 272 AA.
AC OS1878;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP synthase A chain (EC 3.6.3.14) (Protein 6).
GN ATPB.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97361981; PubMed=9216881;
RA Clark M.A., Baumann P.;
RT "The (F1F0) ATP synthase of Buchnera aphidicola (endosymbiont of aphids): genetic analysis of the putative ATP operon.";
RL Curr. Microbiol. 35:84-89(1997).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=98184963; PubMed=9516544;
RA Clark M.A., Baumann L., Baumann P.;
RT "Sequence analysis of a 34.7-kb DNA segment from the genome of Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA, the atp operon, gidA, and rihA.";
RL Curr. Microbiol. 36:158-163(1998).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE (BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), EPSILON(1). CF(0) HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
```

```
EMBL; AF008210; AAC38116.1; -.
DR HSSP; P00855; IC17.
DR InterPro; IPR000568; ATP_synt_A.
DR Pfam; PF00119; ATP_synt_A; 1.
DR PROSITE; PS00449; ATPASE_A; 1.
KW Hydrogen ion transport; Cf(0); Transmembrane.
FT TRANSMEM 41 61
FT TRANSMEM 110 130
FT TRANSMEM 156 176
FT TRANSMEM 188 208
FT TRANSMEM 222 242
FT TRANSMEM 243 263
SQ SEQUENCE 272 AA; 31253 MW; DCD8C7D2C98C37CC CRC64;

Query Match 5.0%; Score 94; DB 1; Length 272;
Best Local Similarity 23.5%; Pred. No. 0.8;
Matches 65; Conservative 45; Mismatches 87; Indels 80; Gaps 16;

QY 73 HYGKVLATVFFYMLVAIIHATIQEYVLDKLSRLQLTKGKQKNEAGQLSVFYIVSG 132
DB 38 HEWLNIDSIFSLVLCGFFLSIF--YTVAK-----KITTVGNPGLQASIELIDFDFIRSN 90
QY 133 IWGMIIASENCLSDP-----TLLWKSQPHNMFTQMKFF-YISQLAYWFHSFPELYFQKV 187
DB 91 VKSM--YQGNPLIAPLSLTVFVWFLNLMDLIPIDFFPFISE--RFFH-FPAM----- 140
```

```
QY 188 RKQDIPGQLIYIGLHFIHGAYLLYLNHLGLLLMLHYAVEL--LSSVCSLLYFGDERY 245
      : : | | | | : | | : : | | : | | : | |
Db 141 --RIVPSADINITLSM-----SLGVFILLIFYSVKMKGLIGFC----- 176
      : : | | | | : | | : : | | : | | : | |
QY 246 QKGLSL-----WPIVFISG---RLVTLIVSVTVGLHLAGTNRGNALSGN----- 288
      : : | | | | : | | : : | | : | | : | |
Db 177 -KELTQPFNHPVFFIFNFLLELVSLSKPISLGLRLF-----GNMYSGEMIFILIAGLL 230
      : : | | | | : | | : : | | : | | : | |
QY 289 -----VNVLAAKIAVLSSSCSTIOVYITWTLTVWL 318
      : | | | : | | : | | : | | : | |
Db 231 PWSQFFLNVPWAIFHL--IISLQAFIFMWLTIVYL 265
      : | | | : | | : | | : | | : | |
```

Search completed: September 6, 2002, 17:21:14
Job time: 632 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:09:51 ; Search time 80.19 seconds
(without alignments)
783.105 Million cell updates/sec

Title: US-09-807-470-2
Perfect score: 1892
Sequence: 1 MGLRKNRNPVLSHEFMV.....NGVENPNRIDSPPKKKEKAP 363

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rotent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Description
1	1125.5	59.5	Q91V04 mus musculus
2	1102	58.2	Q90ZM0 xenopus lae
3	1024	54.1	Q90ZM1 brachydanio
4	762	40.3	Q90ZL9 xenopus lae
5	727	38.4	Q924Z5 mus musculus
6	566	29.9	Q9W5C2 drosophila
7	566	28.9	Q9W5C3 drosophila
8	561	29.7	Q9UIL3 drosophila
9	441.5	23.3	Q9U3P5 caenorhabdi
10	434.5	23.0	Q9XXK7 caenorhabdi
11	348	18.4	Q9CVJ6 mus musculus
12	163.5	8.6	Q9D6J1 mus musculus
13	138.5	7.3	Q9RA82 homo sapien
14	131	6.9	Q9M6A4 lycopersico
15	119.5	6.3	Q95RN6 drosophila
16	116.5	6.2	Q9CHC2 lactococcus

17	115.5	6.1	296	10	Q9LJK3	Q91jk3 arabidopsis
18	115.5	6.1	297	10	Q9M6A2	Q9m6a2 arabidopsis
19	110	5.8	380	4	Q96G23	Q96g23 homo sapien
20	109	5.8	380	11	Q924Z4	Q924z4 mus musculus
21	107.5	5.7	614	16	Q91615	Q91615 bacillus su
22	106.5	5.6	639	16	Q98KL8	Q98kl8 rhizobium l
23	105.5	5.6	310	10	Q97LW0	Q97lw0 arabidopsis
24	103.5	5.4	393	16	Q97LW0	Q97lw0 clostridium
25	103	5.4	414	11	Q9D6K9	Q9d6k9 mus musculus
26	103	5.4	414	11	Q924Z3	Q924z3 mus musculus
27	102.5	5.4	522	16	Q9K9F0	Q9k9f0 bacillus ha
28	101.5	5.4	397	2	O05362	O05362 actinobacil
29	101	5.3	308	10	Q9FZ69	Q9fz69 arabidopsis
30	100.5	5.3	360	5	Q9XWE9	Q9xwe9 caenorhabdi
31	100.5	5.3	397	16	O84825	O84825 chlamydia t
32	100	5.3	354	16	Q9CLS8	Q9cls8 pasteurella
33	100	5.3	406	13	Q90YI6	Q90yi6 brachydanio
34	99.5	5.3	395	16	Q9PLA2	Q9pla2 chlamydia m
35	99	5.2	286	2	Q9X6B4	Q9x6b4 yersinia pe
36	98.5	5.2	441	13	Q91991	Q91991 brachydanio
37	98.5	5.2	460	2	Q9XD16	Q9xd16 leptospira
38	98.5	5.2	722	16	Q9KG28	Q9kg28 bacillus ha
39	98	5.2	232	5	Q9VXF0	Q9vxf0 drosophila
40	97.5	5.2	397	2	O6G245	O6g245 actinobacil
41	96	5.1	378	5	Q9YI17	Q9yi17 halocynthia
42	95.5	5.0	324	2	Q9ZIM4	Q9zim4 listeria mo
43	95.5	5.0	411	17	Q980X8	Q980x8 sulfolobus
44	95.5	5.0	441	13	Q90ZE4	Q90ze4 brachydanio
45	95.5	5.0	751	10	Q93XX4	Q93xx4 arabidopsis

ALIGNMENTS

RESULT 1

Q91V04	PRELIMINARY;	PRT;	374 AA.
AC Q91V04;			
DT 01-DEC-2001 (Tremblrel. 19, Created)			
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE TRAM1 (UNKNOWN) (PROTEIN FOR MGC:11724).			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Hartmann E.;			
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE FROM N.A.			
RA Strausberg R.;			
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AY029764; AAK38167.1; -			
DR EMBL; BC012401; AAH12401.1; -			
SQ SEQUENCE 374 AA; 43039 MW; E6C65250F68E4393 CRC64;			
Query Match	59.5%;	Score 1125.5;	DB 11; Length 374;
Best Local Similarity	58.4%;	Pred. No. 5.6e-93;	
Matches	218;	Conservative 58;	Mismatches 86; Indels 11; Gaps 4;
QY 1 MGLRKNRNPVLSHEFMVQNHDAMVSCVGMFFVLGLMFEGETAEWSIVELTLQHGWWP 60			
Db 1 MATIRKSNKNPPLSHFEFLQNHADIVSCLAMFLGLLMFEVTKAGIAIFVALQYNNTRP 60			
QY 61 A-EGLPGSGRTLYHYGVKDLATVFFYMLVAIIITHATQEVVLKLSRRLQTLTKGKNKLN 119			
Db 61 ATEQATEASLYHYGKDLATVFLYMLVAIIITHATQEVVLKLSRRLQTLTKGKNKLN 120			
QY 120 EAQQLSVFIVSGIWGMIIILASENCLSDPTLLWKSPHNMTFMKFFYISQLAYNFHSF 179			

```
Db 121 ESGLSAFYLFACVNGTIFILISENYISDPTILWRAYPHNLMTFQTKFFYSQLAYWLHAF 180
QY 180 PELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLY 239
Db 181 PELYFQTKKEDIPQLVYIGLYLFIHAGAYLLNHLGLVLLVLYHVEFLFHSRLFY 240
QY 240 FDERYQKGLSWPIVFTISGRVTLIVSVTVGLHLA-GTNRNGNALSGNVNLAAKIAV 298
Db 241 FSDEYQKGFSLWAVFLVGLRLLTLLSVLTGFLARAENQKLDFTGTFNVLAVRIAV 300
QY 299 LSSSCSIQVYITWTLLTWLQRLWLEDANLHVCGRKR-----SRSKGTENGVE---NP 349
Db 301 LASICITQAFMMWKFINQLRRWREHSAFQPPVKRPVAVTKGRSSRKGTENGVTGTSN 360
QY 350 NRIDSPKKKEKA 362
Db 361 NGADSPRNRKEKS 373

RESULT 2
ID Q90ZM0 PRELIMINARY; PRT; 373 AA.
AC Q90ZM0;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TRAM1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029534; AAK40302.1; -.
SQ SEQUENCE 373 AA; 43121 MW; 62AFA9E9859769DF CRC64;
```

```
Query Match 58.2%; Score 1102; DB 13; Length 373;
Best Local Similarity 56.1%; Pred. No. 7.2e-91;
Matches 208; Conservative 64; Mismatches 89; Indels 10; Gaps 3;

QY 1 MGLRKNARNPPVLSHEFWQNHADMVSCVGMFFVLGLMFEGETAEMSIVFTLQHGYYVP 60
Db 1 MGLRKNSTPPVLSHEFIQNHADIVSCVAMVFLGLMFEITAKVAVMFTLQYNTIP 60
QY 61 AGLPSGSRITLY-HYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRLQLTGKONKLINE 120
Db 61 VEGVLGEPTSLHYGIKDMATVFFYMLVAIIHATIQEYVLDKLNRRMHFSKTHSFNE 120
QY 121 AGQLSVFYIVSGIWMIIASENCISDPTILWKSOPHNMTFQMKFFYSQLAYWFHSP 180
Db 121 SQGLSAFYLFSCIGWASIVSENTFSDPISLWKGYPHTFPFQMKFFYSQLAYWFHAF 180
QY 181 ELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLYF 240
Db 181 ELYFQTKKEDIPQLVYIGLYLFIHAGAYLLNHLGLVLLVLYHVEFLFHSRLFYF 240
QY 241 GDERYQKGLSWPIVFTISGRVTLIVSVTVGLHLA-GTNRNGNALSGNVNLAAKIAVL 299
Db 241 SNERYQKFTVMVAVFLVGLRLLTLLSVLTGFLARAENQKLDLSNGNFILAIRITVL 300
QY 300 SSSCSIQVYITWTLLTWLQRLWLEDANLHVCGRKR-----TENGVE---NPN 350
Db 301 ASICITQAFMMWKFINQLRRWREHSSQPSSQKRKATSAKGRKENGVTGTSN 360
QY 351 RIDSPKKKEK 361
Db 361 GADSPRNRKEK 371
```

```
RESULT 3
Q90ZM1
ID Q90ZM1 PRELIMINARY; PRT; 369 AA.
AC Q90ZM1;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TRAM1.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029529; AAK40297.1; -.
SQ SEQUENCE 369 AA; 41743 MW; A3EE6263E3165846 CRC64;
```

```
Query Match 54.1%; Score 1024; DB 13; Length 369;
Best Local Similarity 52.4%; Pred. No. 7.2e-84;
Matches 193; Conservative 67; Mismatches 102; Indels 6; Gaps 4;

QY 1 MGLRKNARNPPVLSHEFWQNHADMVSCVGMFFVLGLMFEGETAEMSIVFTLQHGYYVP 60
Db 1 MGLRKNSTPPVLSHEFIQNHADIVSCVAMVFLGLMFEITAKVAVMFTLQYNTIP 60
QY 61 AGLPSGSRITLY-HYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRLQLTGKONKLINE 119
Db -61 ANGGPEEATVNYFHYGLKDVATIFFYMLVAIIHATIQEYVLDKLNRRMHFSKTHSFNE 120
QY 120 BAGQLSVFYIVSGIWMIIASENCISDPTILWKSOPHNMTFQMKFFYSQLAYWFHSP 179
Db 121 ESGLSAFYLFSCIGWASIVSENTFSDPISLWKGYPHTFPFQMKFFYSQLAYWFHAI 180
QY 180 PELYFQVRKODIPGOLIIYIGLHFIHGAYLLYNHLGLLMLHYAVELLSSVCSLLY 239
Db 181 PELYFQKAKEDIPQLVYISLVLVHAGAYALNLRGLVLLVLYHVEFLFHSRLIY 240
QY 240 FDERYQKGLSWPIVFTISGRVTLIVSVTVGLHLAAGTNRNG-NALSGNVNLAAKIAV 298
Db 241 FSNEERSGTVMVAVFLVGLRLLTLLSVLTGFLAGAEGKQGLNLAEGSNVLFVRVTV 300
QY 299 LSSCSIQVYITWTLLTWLQRLWLEDANLHVCGRK---RRSRKGT-ENGVENPNRIDS 354
Db 301 LAACITQAFMMWKFINQLRRWREQAOTTLKKSSSKSKRANGVNGSVGASGADS 360
QY 355 PPKKKEKA 362
Db 361 PRARKEKS 368
```

```
RESULT 4
Q90ZL9
ID Q90ZL9 PRELIMINARY; PRT; 371 AA.
AC Q90ZL9;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TRAM2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029535; AAK40303.1; -.
SQ SEQUENCE 371 AA; 43548 MW; 582231A5286D130B CRC64;
```

```
Query Match          40.3%; Score 762; DB 13; Length 371;
Best Local Similarity 41.2%; Pred. No. 2.5e-60;
Matches 159; Conservative 67; Mismatches 122; Indels 38; Gaps 7;

QY 1 MGLRKKARNPPVLSHEFMVQNHADMVSCVGMFFVL---GLMFEQTAEMSIIVFLTLQHG 56
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 1 MAFRRR--KSYPLFSQEFVIHNDH---IGFFVLCVLIGLMFEVTAFTAFIFILPQYN 54

QY 57 VVVPABGLPSGRTLYHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRQLTKGKQN 116
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 55 SSIQT---LDGEILYHYGVKDLVTLFVWIAIIHLAIHQEYILDINKRLHLSKVQS 111

QY 117 KLNAGOLSVFYIVSGIWMGMIILASENCLSDPTLLWKSSOPHNMFTOMQFFYSIQLAYWF 176
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 112 RENESQLAFLHASMFWCLYSATSEGYLSYPKTLMESYPHYLPQVFFVLCQAYWL 171

QY 177 HSPPELYFOKVRKQDIPGOLIYIGLHFHIGGAYLLYLNHLGILLMLHYAVELLSSVCS 236
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 172 HALPELYFOKVRKEEVPRLQYIVLYLLHAGAYLLNLTLGLILLLLQSVAEFLFHAR 231

QY 237 LLYFGDERYQKGLSMPVIFISGRLVTLIVSVVTVGLHLAGYN-RNGNALSGNVNVLAAK 295
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 232 LFYFTDENNQRLFNAGVGVFVITRLFTLTLVSLVITGFLARAEEVHTFPDKGTINTLLFR 291

QY 296 IAVLSSCSIQVYITWTLTTVWLORWLEDANLHVCGKRRSRKGT----- 342
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 292 MVVLLMCVSQTMWMMRFTHFQLRRWRE-----CCKEQAAKRRSVAVAMMKQAKVIKR 345

QY 343 -----ENGVENPNRIDSPPKKKEKAP 363
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 346 ESSYHENGVVKAENGSTPQKKIKSP 371

RESULT 5
Q92425 ID Q92425 PRELIMINARY; PRT; 370 AA.
AC Q92425;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE EG:BACR7A4.5 OR CG11642.
GN EG:BACR7A4.5 OR CG11642.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroides; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceinikher S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Balgwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fodor C.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirsaks R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhou S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003419; AAF45569.1; -.
DR FlyBase; FBgn0040340; EG:BACR7A4.5.
```

```
Query Match          38.4%; Score 727; DB 11; Length 370;
Best Local Similarity 40.7%; Pred. No. 3.4e-57;
Matches 151; Conservative 70; Mismatches 134; Indels 16; Gaps 5;

QY 5 KKNARNPPVLSHEFMVQNHADMVSCVGMFFVLGLMFEQTAEMSIIVFLTLQHGVPVPAEGL 64
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 4 RRRRTSYPLFSQEFIIHNDHADIGFCLVLCVLIGLMFEVTAFTAFILPQYNISVPT--- 60

QY 65 PSGSRTL-YHYGVKDLATVFFYMLVAIIHATIQEYVLDKLSRRQLTKGKKNLEACQ 123
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 61 -ADSETVHYHYGVKDLVTLFVWVITIFHVAVYILDKISKRHLHLSKVSKFNESQ 119

QY 124 LSVFYIVSGIWMGMIILASENCLSDPTLLWKSSOPHNMFTOMQFFYSIQLAYWFHSPPELY 183
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 120 LLVFHL SAVAWCFYIVTEGYITNPSLWEDYPHYLSFQVAFYILGQAYWLHLSPELY 179

QY 184 FOKVRKQDIPGOLIYIGLHFHIGGAYLLYLNHLGILLMLHYAVELLSSVCSLLYFGDE 243
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
```



```
Db 67 AVEQOQEREVHGVLGSLDPAIFVYVCWIVHVAQVQGLDKSKKTHLSKVSTFRFG 126
QY 120 EAQSLVFEYIVSGWGMILASENCISD-----PILLWKSQP--HNMFTQMKFFYISQL 172
Db 127 ESFH-QMFTVTSIAHAFYISER-LEDFSEYKSFVSVMLGYPTERHVSAAKLYFIQI 184
QY 173 AYWFHSFPPELYPQKVRKODIPQOLYIGLHFLPHGIGGAYLLYVNLHGLLLMLHVAVELLS 232
Db 185 SYWIHQFPEFYQLKRDREIROKSVQAILHHTAFISIAFFNFTRVGLALITILEYITQIF 244
QY 233 SVCSLLYFGDERYQKGLS-----LMPVIFISGRVLTIVSVVT--VGLHLAGTNRNGNAL 285
Db 245 HIARFAHVG---RKLSDPAFKLFGSVFLVRLGSLIIIAVMTFWYGLRQA-ESPFVDIS 300
QY 286 SGNVNVLAAKIAVSSSCSIQVITLTTVTLWLRQWLEDANLHVCGRKRRSRKRGTENG 345
Db 301 AGNFTAVIRLVNLAALLVLLQFLYISPV-----FHM-GRFRESNAKREKKS 348
QY 346 VENPNRIDSPPKKKEK 361
Db 349 AAAAAAV---PKKEK 361

RESULT 11
Q9CVJ6 PRELIMINARY; PRT; 159 AA.
AC Q9CVJ6;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE 1810049E02RIK PROTEIN (FRAGMENT).
GN 1810049E02RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK013554; BAB28903.1; -
DR EMBL; BC003946; AAH03946.1; -
DR EMBL; AY029531; AAK40299.1; -
DR MGD; MGI:1914510; 2900019C14Rik.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 159 AA; 18053 MW; D88C0B3126B0085A CRC64;
```

```
Query Match 18.4%; Score 348; DB 11; Length 159;
Best Local Similarity 48.1%; Pred. No. 1.4e-23;
Matches 76; Conservative 22; Mismatches 50; Indels 10; Gaps 3;
QY 215 NHLGLLLMLHYAVELLSSVCSLLYFGDERYQKGLSLWPTVIFISGRVLTIVSVVTGLH 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
Db 1 NHLGLVLLVLYHYVEFLFHISRLFEVSDKYQKGLSLWAVLFLVGRLLTLILSVLTGFG 60
QY 275 LA-GTNRGNALSNVNVLAAKIAVSSSCSIQVITLTTVTLWLRQWLEDANLHVCGRK 333
Db 61 LARAENOKLDFSTGNFNVLAIRIAVLAISICITAFMMWKEINFQLRWRREHSFAQAPVK 120
QY 334 RR-----SRSRKGTENGVE---NPNRIDSPPKKKEKA 362
Db 121 RKPAVTKGRSSRKGTENGVCNCTVTSNGADSPNRKES 158

RESULT 12
Q9D6J1 PRELIMINARY; PRT; 393 AA.
AC Q9D6J1;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (ITEMBLrel. 19, Last annotation update)
DE 2900019C14RIK PROTEIN (RIKEN CDNA 2900019C14 GENE) (TRH1).
GN 2900019C14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK013554; BAB28903.1; -
DR EMBL; BC003946; AAH03946.1; -
DR EMBL; AY029531; AAK40299.1; -
DR MGD; MGI:1914510; 2900019C14Rik.
DR InterPro; IPR001356; Homeobox.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
SQ SEQUENCE 393 AA; 46016 MW; B2038F7A128F816F CRC64;
```

```
Query Match 8.6%; Score 163.5; DB 11; Length 393;
Best Local Similarity 26.0%; Pred. No. 1.5e-06;
Matches 46; Conservative 40; Mismatches 76; Indels 15; Gaps 5;
QY 106 RRLQLTGKQNKLINEAGQLSVFYIVSGWGMILASENCISDPTLLWKSQPHNMFTQMK 165
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 RRNQDRPSLSKKFCEACWRVFEVYLCFSVGTSTLYHESWLWSLWSPHWTQNLNLSLY 180
```

GN	ASC.
OS	Lycopersicon esculentum (Tomato).
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX	NCBI_TaxID=4081;
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=CV_VFNT CHERRY;
RC	MEDLINE=99168767; PubMed=10071209;
RX	Mesbah L.A., Kneppers T.J., Takken F.L., Laurent P., Hille J.,
RA	Nijkamp H.J.;
RA	"Genetic and physical analysis of a YAC contig spanning the fungal
RT	disease resistance locus Asc of tomato (Lycopersicon esculentum).";
RL	Mol. Gen. Genet. 261:50-57(1999).
[2]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=CV_VFNT CHERRY;
RC	MEDLINE=20243803; PubMed=10781105;
RX	Brandwagt B.F., Mesbah L.A., Takken F.L.W., Laurent P.L.,
RA	Kneppers T.J.A., Hille J., Nijkamp H.J.J.;
RA	"A longevity assurance gene homolog of tomato mediates resistance to
RT	Alternaria alternata f. sp. lycopersici toxins and fumonisin Bl.";
RL	Proc. Natl. Acad. Sci. U.S.A. 97:4961-4966(2000).
DR	EMBL; AF198177; AAP67518.1; -.
SQ	SEQUENCE 308 AA; 36301 MW; 2BD3378CE53F416D CRC64;
Query Match	6.9%; Score 131; DB 10; Length 308;
Best Local Similarity	19.5%; Pred. No. 0.00094;
Matches 46; Conservative 50; Mismatches 80; Indels 60; Gaps	
QY	62 EGLPGSGRTLHYGVKDLATVFYMLVAIIHATHIOEYVLDKLSRRLOLTK-----L126
DB	16 ESUPE-----YQDLIFELLFAFFPVLRIILDFVEALAKRMIFGCKTVVNINGR 66
	: : : : : : : : : : : : : : : :
QY	113 ---GKONKLNAGQLSVFYIVSGIWGMIIILASCNCLSDPTLLMKSQ-----PHNMFTQM 164
DB	67 EERKKINKPFESAWEKVFYFLSAELLALSVCNEPWFTDSRYFWAGPGDVVPNLKMKLK 126
	: : : : : : : : : : : : : : : :
QY	165 KPFIYSOLAWFHIS-PPELYFKVRKODIPGOLIYYGLHLFHFGAYLYLVNH-----L 217
DB	127 KLLYMAGGGFYFSIFATLYWE--TRRYDFAAQIHH-----HVTTSVLIVSYVYGfARI 179
	: : : : : : : : : : : : : : : : : :
QY	218 GLLLLMLHYAVELSSVCSL-----LYFGDERYQKGLSLWPPIVFISGRLV 262
DB	180 GSVVALHDGSDVFMEIAKMSKSGGFDLIADIFF-----SLFALVFTSLRII 226
	: : : : : : : : : : : : : : : : : :
RESULT 15	
Q95RN6	PRELIMINARY; PRT; 400 AA.
ID Q95RN6	
AC Q95RN6;	
DT 01-DRC-2001 (TrEMBLrel. 19, Created)	
DT 01-DRC-2001 (TrEMBLrel. 19, Last sequence update)	
DT 01-DRC-2001 (TrEMBLrel. 19, Last annotation update)	
DE LD18904P.	
GS C61S898.	
OS Drosophila melanogaster (Fruit fly).	
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC Ephydroidea; Drosophilidae; Drosophila.	
OX NCBI_TaxID=7227;	
[1]	
RN SEQUENCE FROM N.A.	
RP STRAIN=Y, CN BW SP.	
RC Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,	
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,	
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,	
RA Nunco J., Pacble J., Paragas V., Park S., Phouanenavong S., Wan K.,	
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;	
RL Submitted (EMBL-2001) to the EMBL/GenBank/DDJB databases.	
RL EMBL; AY061255; AAL28803.1; -.	

SQ SEQUENCE 400 AA; 46351 MW; AB586DB2B811E62 CRC64;

Query Match 6.3%; Score 119.5; DB 5; Length 400;
Best Local Similarity 18.8%; Pred. No. 0.014;
Matches 49; Conservative 52; Mismatches 96; Indels 63; Gaps 7;

QY 61 AEGLPSSRTLYHVGKDLATVFYMLVATIIHATIOEY----- 99
Db 25 ADIAPGSRPDVVHANYKDLIWPFPFAAVMLVRYTLERFWISPYGKSLGTRSRPKKAA 84

QY 100 --VLDK--LSRRLQLTGKQKNLNEAGOLS-----V 126
Db 85 VPILEKTYAKSTRLDKKKLVPLSKQTDMSEREIERWRLRRAQDKPSTLVKFCENTWRCI 144

QY 127 FYIVSGIWMIIASENCLSDPTLLWKSQPHNMWTFOMKEFY-ISQLAYWFHSFPPELYFQ 185
Db 145 YLYSFIYFGVIVLWDKPWFVDKSCWYGYPHQSISNDIWWYIMISFYW--SLTGTQFF 202

QY 186 KVRKQDIPGOLIVIGLHFIHGAYLLYLNHLGILLMLHYAVELLSSVCSSLYFGDERY 245
Db 203 DVKKDFQWQFIHWMVTLMLSLWSVCNHRVGSVLVVDHCADIFLEAAKLTXYA--KY 260

QY 246 QK---GLSLWPIVVISGRL 261
Db 261 QKLCDAIFAIFTVVWIVTRL 280

Search completed: September 6, 2002, 17:20:26
Job time: 635 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 14:46:34 ; Search time 5654.61 Seconds
(without alignments)
8467.413 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcagcggtgctga.....ccacttgtaaatgggaattc 2288

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_bt.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description

c	1	2254.8	98.5	165538	9	AC093805	AC093805 Homo sapi
2	1288	56.3	1288	6	BD005228	BD005228 Method of	
3	1110	48.5	1110	6	BD005226	BD005226 Method of	
4	961.8	42.0	2856	4	CFTRAM	X63678 C.familiari	
5	937.6	41.0	2722	9	BC000687	BC000687 Homo sapi	
6	804.8	35.2	2387	4	BTU19578	U19578 Bos taurus	
7	803.8	35.1	155623	2	AC106086	AC106086 Rattus no	
8	742.4	32.4	1267	6	AX329795	AX329795 Sequence	
9	742.4	32.4	1267	6	BD005229	BD005229 Method of	
10	742.4	32.4	1267	9	HSTRAMP	X63679 H.sapiens m	
11	733	32.0	2720	10	AY029764	AY029764 Mus muscu	
12	733	32.0	2819	10	BC012401	BC012401 Mus muscu	
13	648.2	28.3	73339	2	AC027246	AC027246 Homo sapi	
14	647.8	28.3	1092	6	BD005227	BD005227 Method of	
15	508.8	22.2	1338	5	AY029534	AY029534 Xenopus l	
16	486.8	21.3	642	6	AX303127	AX303127 Sequence	
17	443.8	19.4	1453	5	AY029529	AY029529 Danio rer	
18	297.8	13.0	6974	9	HUMORFKG1A	D31762 Human mRNA	
19	269	11.8	1360	10	AY029530	AY029530 Mus muscu	
20	269	11.8	1408	10	BC018212	BC018212 Mus muscu	
21	258.4	11.3	1272	5	AY029535	AY029535 Xenopus l	
22	233.4	10.2	196349	2	AC022032	AC022032 Homo sapi	
23	231.8	10.1	163997	9	AC022731	AC022731 Homo sapi	
24	161	7.0	2123	3	AY075393	AY075393 Drosophil	
25	161	7.0	132910	2	AC014319	AC014319 Drosophil	
26	161	7.0	146717	3	AC097726	AC097726 Drosophil	
27	161	7.0	212370	3	DMBR7A4	AL109630 Drosophil	
28	161	7.0	299275	3	AE003419	AE003419 Drosophil	
29	118	5.2	447	6	AX192932	AX192932 Sequence	
30	118	5.2	196349	2	AC022032	AC022032 Homo sapi	
31	94.8	4.1	64530	2	AC110042	AC110042 Mus muscu	
32	93.8	4.1	189572	2	AC084812	AC084812 Homo sapi	
33	88.4	3.9	494	6	AX305631	AX305631 Sequence	
34	87.2	3.8	41805	3	CBRG44C02	AC084610 Caenorhab	
35	81	3.5	420	11	G37645	G37645 SHGC-38238	
36	69.8	3.1	1017	9	AF274933	AF274933 Homo sapi	
37	65.6	2.9	602	6	AX285010	AX285010 Sequence	
38	64.2	2.8	189572	2	AC084812	AC084812 Homo sapi	
39	63.2	2.8	109813	9	HSJ180E22	AL049611 Human DNA	
40	62.2	2.7	14544	3	CEC24F3	AL022716 Caenorhab	
41	60.8	2.7	2468	8	YSCMTORIK	M11449 Yeast mitoc	
42	60.8	2.7	21153	8	YSCMTGGL3	L36897 Saccharomyc	
43	59.4	2.6	155973	2	AC100727	AC100727 Mus muscu	
44	58.2	2.5	7218	6	I66494	I66494 Sequence 14	
45	57.2	2.5	97672	8	AC006951	AC006951 Arabidops	

ALIGNMENTS

RESULT	1	AC093805/c	165538 bp	DNA	linear	PRI 08-FEB-2002
AC093805/c	Homo sapiens BAC clone RP11-32609 from 4, complete sequence.					
LOCUS	AC093805	AC012532				
DEFINITION	AC093805	AC012532				
ACCESSION	AC093805.3	GI:18497223				
VERSION	HTG.					
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 165538)					
TITLE	Toward a complete human genome sequence					
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)					
MEDLINE	99063792					
REFERENCE	2 (bases 1 to 165538)					
AUTHORS	Cedroni, M., Abbott, A. and Bielicki, L.					
TITLE	The sequence of Homo sapiens BAC clone RP11-32609					
JOURNAL	Unpublished (2001)					
REFERENCE	3 (bases 1 to 165538)					
AUTHORS	Waterston, R.H.					
TITLE	Direct Submission					

JOURNAL

Submitted (10-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

4 (bases 1 to 165538)

REFERENCE

AUTHORS

Waterston,R.H.

Direct Submission

JOURNAL

Submitted (05-FEB-2002) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

5 (bases 1 to 165538)

REFERENCE

AUTHORS

Waterston,R.

Direct Submission

JOURNAL

Submitted (08-FEB-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Feb 5, 2002 this sequence version replaced gi:16077043.

COMMENT

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0326009
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-32609;
actual end is at base position 165538 of RP11-32609.

Data from AC013556 was used to finish the clone, AC093805.

Polymorphisms have been identified between AC013556 and AC093805.

The sequence of AC012532 has been incorporated into AC093805.

FEATURES

source

Location/Qualifiers

1. .165538

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="4"

/map="4"

/clone="RP11-32609"

/clone_lib="RPCI-11"

1. .130

/rpt_family="MaLR"

repeat_region

repeat_region 131. .436
/rpt_family="ERV1"
repeat_region 870. .1245
/rpt_family="L1"
repeat_region 1246. .2720
/rpt_family="L1"
repeat_region 2721. .3067
/rpt_family="L1"
repeat_region 3627. .3713
/rpt_family="MaLR"
repeat_region 4217. .5074
/rpt_family="Achoho"
repeat_region 5055. .5101
/rpt_family="(TTTA)n"
repeat_region 5075. .5384
/rpt_family="Alu"
repeat_region 5385. .5592
/rpt_family="Achoho"
repeat_region 7152. .7426
/rpt_family="L1"
repeat_region 7427. .7734
/rpt_family="Alu"
repeat_region 7590. .7619
/rpt_family="AT_rich"
repeat_region 7735. .7872
/rpt_family="L1"
repeat_region 7873. .8314
/rpt_family="ERV1"
repeat_region 8315. .9564
/rpt_family="L1"
repeat_region 9438. .9470
/rpt_family="(A)n"
repeat_region 9822. .9853
/rpt_family="(CA)n"
repeat_region 10274. .10722
/rpt_family="MaLR"
repeat_region 11045. .11163
/rpt_family="CT-rich"
repeat_region 13136. .13287
/rpt_family="Alu"
repeat_region 15545. .15567
/rpt_family="AT_rich"
repeat_region 15850. .15875
/rpt_family="AT_rich"
repeat_region 15929. .16234
/rpt_family="Alu"
repeat_region 17161. .17194
/rpt_family="AT_rich"
repeat_region 18540. .18828
/rpt_family="MaLR"
repeat_region 18938. .19023
/rpt_family="MaLR"
repeat_region 19181. .19224
/rpt_family="MaLR"
repeat_region 19256. .19588
/rpt_family="MaLR"
repeat_region 19659. .19726
/rpt_family="MaLR"
repeat_region 19847. .20312
/rpt_family="MER2_type"
repeat_region 19928. .19960
/rpt_family="(TG)n"
repeat_region 20513. .20623
/rpt_family="Alu"
repeat_region 20596. .20640
/rpt_family="(A)n"
repeat_region 20779. .20802
/rpt_family="AT_rich"
repeat_region 20868. .21156
/rpt_family="Alu"
repeat_region 21487. .21514
/rpt_family="AT_rich"
repeat_region 21958. .21983

Qy	1651	agagatttttctgttttgaaatagtcgtctcttcggtttttttgtattgaactgttt	1710
Db	47780	AGAGATTTTCTGTTTGTGAAATAGTTCGTGCTCTTCGGTGTGTTTGTATTGAACCTGTT	47721
Qy	1711	catgtatttttaagacatttgaagggaagagattattatgaatgggaaaaaagatt	1770
Db	47720	CATGTATTTTAAAGACATTTGAGGGGAGGAGGATTATTATGAATGGGAAAAAAGATT	47661
Qy	1771	ttggtgagactaaattactcatcgtaaaataatgcaaaaatagtttttgggatacaca	1830
Db	47660	TTGGTTGAGACTAAATPACTCATCGTCAAAATAATGTCAAAATAGTTTGGGGATCACCA	47601
Qy	1831	ctatatttgttttgatttttaacotttcaacattttccctaatttccctaagatttcagagataact	1890
Db	47600	CTATATTTTGTGTTTGAATTTTAACTTTCAACATTTTCCTAATGATTTGCAGAGATAACT	47541
Qy	1891	gcacaatttgcatacataatgatactagggttcttactccccaccagtggtttcataatactaa	1950
Db	47540	GCACAAATTTTGCAATPACATGATACTGGTTCCTTACTCTCCACCACTGTTTCATAATACTAA	47481
Qy	1951	caagatggtctctctagcaagattatgttttaagtcttgcttttggggtaaaaaaag	2010
Db	47480	CAAGATGGTCTCCTAGCAAGATTATGTGTTAATGCTTGTTCCTTTGGGGTAAAAATAAAG	47421
Qy	2011	tacgaaaagtggaagtaaaatcagtatctgtaattgttagaatttatttttaagaa	2070
Db	47420	TACGAAAAGGTGGAAGTCAAAATCAGTATTCGTGAATGTTAGATTTATTTTAAAGAA	47361
Qy	2071	cttaactcagaaaaagattgtagactcaccaaaaataataatgttctttattttacag	2130
Db	47360	CTTACAACTCAGAAAAGATTGCTAGACTCACCAAAATAATAAATGTTCTTTATTTTACAG	47301
Qy	2131	gtagtattattagtgcttcattccccatttaaaaaaacacagtaactaatgggtaacacat	2190
Db	47300	GTAGTGATTATTAGTGCTTCATCCCCAFTTTAAAAAACACAGTACTAATGGGTAAACACAT	47241
Qy	2191	atggaggttgctgccatatatttgatcaaaaatacattaatataataataataata	2250
Db	47240	ATGGAGGTTTGCTGCCATATATATTGCAATCAAAATATCATTAATATATAAANAATTA	47181
Qy	2251	aaatcatctcgtccatccacttgtaaaatgggaattc	2288
Db	47180	AAATCATCTGTGTCATTCACCTGTAATGGGAATTC	47143
RESULT 2			
LOCUS	BD005228	1288 bp	DNA linear PAT 31-JAN-2002
DEFINITION	Method of screening of protein.		
ACCESSION	BD005228		
VERSION	BD005228.1	GI:18633189	
KEYWORDS	JP 03075332-T/3.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1288)		
TITLE	Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.		
JOURNAL	Method of screening of protein Patent: JP 03075332-T 3 16-FEB-2001; SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA, OTOAKI IMAMURA, HIROKORI ISHIKAWA, KIYOMITSU NEMOTO		
COMMENT	OS Homo sapiens (human)		
	PN	JP 03075332-T/3	
	PD	16-FEB-2001	
	PF	17-AUG-2000 JP 2000005488	
	PR	20-AUG-1999 JP 99P 234764	
	PI	NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIROKORI ISHIKAWA, KIYOMITSU NEMOTO	
	PC	C12Q1/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC (C12P21/02,C12R1:91)	
CC			
PH	Key	Location/Qualifiers	

FEATURES	FT	CDS	(40)..(1146).	
source		Location/Qualifiers	1..1288	
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
BASE COUNT	331 a	275 c	276 g	406 t
ORIGIN				
	Query Match	56.3%	Score 1288;	DB 6; Length 1288;
	Best Local Similarity	100.0%;	Pred. No. 1.1e-300;	
	Matches 1288;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	438	cagctgcgcgtcgcggtgagtatccgggaagcgccaccatggggctccgtaagaagacg	497	
Db	1	CAGCTGCGGCTCGCGGTGAGTATCGGGGAAGCCACCATGGGGCTCCGTAAAGAGAC	60	
Qy	498	accaagaaccccccggttctcagccaggaaattcatctcgcagaatcatcgggacatcgc	557	
Db	61	ACCAAGAACCCTCCGCTTCTCAGCCAGGAATTCATCTGCAGAAATCATGCGGACATCGTC	120	
Qy	558	tcctgcgtgggagatgttcttctcgtcgtgggcttgcgttcgagggaaacagcagaagatcc	617	
Db	121	TCCTGCGTGGGGATGTCTTCCTGCTGGGGCTTGTGTTTCAGGGAACAGCAGAAAGCATCC	180	
Qy	618	atcgtgttctcactcttcagcacagtgctgcgtccctgcagcagaggaacaaagccacg	677	
Db	181	ATCGTGTTCCTACTCTTCAGCACAGTGTGCTGCTCCTGCAGCAGAGGAACAAAGCCACG	240	
Qy	678	ggctcaaaagtcctctctattattggtgtcaaaagatttggccacggtttctctctacatg	737	
Db	241	GGCTCAAAGTCCCTCTATTATTATGTTGTCAAAGATTGGCCACGGTTTCTTCTACATG	300	
Qy	738	ctggtggcaatcattatcgcacaaattcaggaatatgtgttggataaaataaacaag	797	
Db	301	CTGGTGGCAATCATATTTCATGCCACAATTCAGGAATATGTGTGGATAAATAAACAAG	360	
Qy	798	agaatcagttccaaagcgaacaaacaaagtttaacagctcgttcagtttagtggtg	857	
Db	361	AGAAATGCAGTTCACCAAGCGAAGCAAAACAAAGTTTAAACAGTCTGGTCAAGTTAGTGTG	420	
Qy	858	ttctacttttttctgtatttggggcacattcttaattctctgaaactcacttccagaaa	917	
Db	421	TTCTACTTTTTTCTGTGATTTGGGCACATTTCATTTAATCTCTGAAAACCTGCTGTCA	480	
Qy	918	gacccaactcttatagggaaggctcgtccccatagcatgatgacatttcaaaatgaagttt	977	
Db	481	GACCCCAACTCTTTATATGGAAGGCTCGTCCCATAGCATGATGACATTTCAAAATGAAGTTT	540	
Qy	978	ttctacatatcccgatggccttactggttttcagctgttcttctgaaactctacttccagaaa	1037	
Db	541	TTCTACATATCCCAAGTTGGCTTACTGGTTTCATGCTTTTCTCTGAACCTACTTCCAGAAA	600	
Qy	1038	accaaaaacaagacatccctcgtcaactgtctacattgtgtcttccactcttccacatt	1097	
Db	601	ACCAAAAAACAAGACATCCCTCGTCAACTTGTCTACATTTGGTCTTCACCTTCCACATT	660	
Qy	1098	actggagcttatctctgtacttgaatcatttggggaactcttcttcttggtagctgattat	1157	
Db	661	ACTGAGCTTATCTCTGTACTTGAATCATTTGGGACTTCTCTCTTTTGGTACTGCATTAT	720	
Qy	1158	tttgtgaattactttcccacatgctgcgcgtgttttacttttagttagtgaagatccag	1217	
Db	721	TTTGTGAATTACTTTCCACATGTCGGGCTGTTTTTACTTTTACTGATGAAAAGTACCAG	780	
Qy	1218	aaaggcatatctctgtgggccattgttttatctctgtggtagacttgtgactttaattgtt	1277	
Db	781	AAAGGCATATCTCTGTGGGCATTTGTGTTTATCTTGGTAGACTTGTGACTTTAATTTGTT	840	
Qy	1278	tccttactcaactgttgggtttcaactggctggatgcgagaatcgggaatcctgatgcctt	1337	
Db	841	TCCGTACTCAGTGTGGGTTTCCCTGGCTGGATCGAGAATCGGAATCTCTGTATGCCCTT	900	

```
QY 1338 actggaatgtaaatgttggtggcagctaaattgctgttctgtcgtccagtgtaacagatc 1397
Db 901 ACTGGAAATGTAATGTGTGGCAGCTAAATTTGCTGTTCTGTCGTCAGTTGCACGATC 960
QY 1398 caagcctaagtaacatggaaacttaatactctctgtggtctcagaggtggttagaagattct 1457
Db 961 CAAGCCTACGTAACTGGAACCTAAATTTACTCTCTGCTTCCAGAGTGGGTAGAAGATTCT 1020
QY 1458 aatattcagcctcatgtatgaaagaaacagctcagatcttcttaaaaaaagaaacagaa 1517
Db 1021 AATATTTCAGCCCTCATGTATGAAAGAAACAGCTGAGATCTTCTAAAAAAGAAACAGAA 1080
QY 1518 aacggagtggagtggaacttcaaatagatgagctgcccgaagaggaagaaagagaa 1577
Db 1081 AACGGAGTGGAGTGGAAACTTCAATAGATGAGACTGCCGCAAGAGGAAGAGAGAA 1140
QY 1578 tcttcataatctttgcaagcgaattgatttaattctctgcaaaagaactcgtctctttgaggt 1637
Db 1141 TCTTCATAATCTTTGCAAGCGCATGATTAAATGCTGCAAAAGGAATCTGCTCTTTGAGGT 1200
QY 1638 ttctttctgactagagattttctgtttttgaaaaatagttcgtctcttcggtttttgt 1697
Db 1201 TTCTTTCTGCACGTAGAGATTTCCTGTTTGTGAAATAGTTCGTGCTCTTCGTTTTTGT 1260
QY 1698 tattgaactgtttcatgtatttttaa 1725
Db 1261 TATTGAACGTGTTTCATGTATTTTTAAA 1288

RESULT 3
BD005226
LOCUS BD005226 1110 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005226
VERSION BD005226.1 GI:18633187
KEYWORDS JP 03075332-r/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1110)
Todo.N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-r 1 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD, NAOKI TODO, HAJIME OKUYAMA, OTOAKI
IMAMURA, HIROKORI ISHIKAWA, KIYOMITSU NEMOTO
OS Homo sapiens (human)
PN JP 03075332-r/1
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PR 20-AUG-1999 JP 99P 234764
PI NAOKI TODO, HAJIME OKUYAMA, MOTOAKI IMAMURA, HIROKORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C12Q1/02.G01N33/50.C07K14/47.A61K38/17.C12N5/10.C12P21/02// PC
(C12P21/02.C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1110
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 296 a 241 c 236 g 337 t
ORIGIN

Query Match 48.5%; Score 1110; DB 6; Length 1110;
Best Local Similarity 100.0%; Pred. No. 1.2e-257;
Matches 1110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 atggggctcctaagaagacacaaagaccccccttctcagccaggaattcatcctg 536
Db 1 ATGGGGCTCCGTAAGAAGAGCACCAGAACCCCGTCTTCAGCCAGGAATTCATCTCTG 60
```

```
QY 537 cagaatcatcgcgacatcgtctcctcgtggggatgtctctcctgctggggcttggttc 596
Db 61 CAGAAATCATCGGACATCGTCTCCTCGTGGGATGTTCTTCTGCTGGGGCTTGTGTTT 120
QY 597 gagggaaacagcaagaacatccatcgtgtttctcactcttcagcacagtggtgtgctccct 656
Db 121 GAGGGAACAGCAGCAAGATCCATCGTGTCTTCTCCTTCTCAGCACAGTGTGTGCTGCCCT 180
QY 657 gcagcagaagaaacacccagcgggtccaaagtccctctattattatggtgcaagatttg 716
Db 181 GCAGCAGAGAAACAAAGCCAGGGCTCAAGTCCCTCTATTATTATGTTGTCGAAGATTTG 240
QY 717 gccacggtttctctacatgctgtggcaatcatctattcatcccaatcagggaatat 776
Db 241 GCCACGGTTTCTTCTTACATGCTGGTGGCAATCATATTATCATGCCACAATTCAGGAATAT 300
QY 777 gttgtgataaaattaaacagagaatgcagttccacaaagcgaacaaacaaagttaac 836
Db 301 GTGTTGGATAAAATTAACAAGAGAATGCAGTTTCCACCAAGCGAAACAAAACAAAGTTTAA 360
QY 837 gagctcgtcagtttagtggttctacttttttttttttttttttttttttttttttttt 896
Db 361 GAGTCTGGGTGAGTTAGTGTGTCTTCTTTTCTTGTATTGTTGGGCACATTCATTTTA 420
QY 897 atctctgaaaactgctgtcagaccacaaactcttatatggaagctcgtcgtcccatagcatg 956
Db 421 ATCTCTGAAAACATGCTGTGTCAGACCCCACTTATATGGAAGGCTCGTCCCATAGCATG 480
QY 957 atgacatttcaaatgaagttttctacatatcccgcttggtggttctcattgctctttt 1016
Db 481 ATGACATTTCAAAATGAAGTTTCTTCTACATATCCAGTTGGCTTACTGGTTTCATGCTTTT 540
QY 1017 cctgaactctactcagaaaaacaaaaacaaagacatccctcgtcaactgctcaactt 1076
Db 541 CCTGAACCTCTACTTCCAGAAAACAAAACAAAGACATCCCTCGTCAACTGTCTACATT 600
QY 1077 ggtcttcacactctccacattactgaggttatctctgttacttgaaatcatttgggactt 1136
Db 601 GGTCTTCACTCTTCCACATTACTGGAGCTTATCTCTGTACTTGAATCATTTGGGACTT 660
QY 1137 cttcttttggtactgcatattttgttgaaataactttccacatgctgctgctgttttac 1196
Db 661 CTTCTTTTGGTACTGCAATATTATTTTGTGAATTAATTTCCACATGTGCGGCCCTGTTTAC 720
QY 1197 tttagtgatgaaagaccagaaagcatactctcgtgggccaattggttttactcttggtt 1256
Db 721 TTTAGTGTGATAAAGTACCAGAAAAGGCATATCTCTGTGGGCCATTGTGTTTATCTTGGGT 780
QY 1257 agactgtgactttaattgtttccgtactcaactgttgggtttcacactggctggatcgacg 1316
Db 781 AGACTTGTGACTTAAATTTGTTCCGTACTCCTGTTGGGTTCACCTGGCTGGATCGCAG 840
QY 1317 aatcggaactcctgatgcccttactggaaatgtaaatgtgttggcagcgttaaaattgctgt 1376
Db 841 AATCGGAATCCTGATGCCCTTACTTGGAAATGTAAATGTGTGTGGCAGTAAATTTGCTGTT 900
QY 1377 ctgtcgtcagttgcaagctcagcgttaacatggaacttaattactctcctgctt 1436
Db 901 CTGTCTGCCAGTTGCAGATCCAGGCTACGTAACTGGAACCTTAAATTAATCTCTGCTTGGCTT 960
QY 1437 cagagtggttagaagattcttaatttcagggcctcatgtatgaaaaagaaacggtcgaga 1496
Db 961 CAGAGGTGGGTAGAAGATTTCTTAATATTCAGGCCCTCATGTATGAAAAGAAACGGTCGAGA 1020
QY 1497 tcttctaaaaaagaaacagaaacggagtggtggagtggaacttcaaatagtagactgt 1556
Db 1021 TCTTCTAAAAAAGAACAGAAACGGAGTGGGAGTGGAAACTTCAAAATAGAGTAGACTGT 1080
QY 1557 cgcctaaagaggaaagaaatcttcaataa 1586
Db 1081 CCGCCAAAGAGGAAAGAGAAATCTTCATAA 1110
```

```
RESULT 4
CFTRAM
LOCUS C.familiaris mRNA for TRAM-protein. 2856 bp mRNA linear MAM 02-JUN-1992
DEFINITION X63678.1
ACCESSION X63678.1 GI:941
VERSION 1
KEYWORDS TRAM-protein.
SOURCE dog.
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 2856)
AUTHORS Gorlich,D., Hartmann,E., Prehn,S. and Rapoport,T.A.
TITLE A protein of the endoplasmic reticulum involved early in
polypeptide translocation
JOURNAL Nature 357 (6373), 47-52 (1992)
MEDLINE 92244357
REFERENCE 2 (bases 1 to 2856)
AUTHORS Hartmann,E.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1992) E. Hartmann, Max-Delbrueck-Centr. f.
Molekulare Med., Robert-Rössle-Strasse 10, O-1115 Berlin Buch, FRG
FEATURES
source
1..2856
/organism="canis familiaris"
/db_xref="taxon:9615"
/cell_line="MDCK"
219..1343
/codon_start=1
/product="TRAM-protein"
/protein_id="CAA45217.1"
/db_xref="GI:942"
/db_xref="SWISS-PROT:Q01685"
/translation="MAIRKSKTKSPVLVLSHEFILONHADIVSCVAMVFLGLMPEITA
KASLIFVTLQNVNLPATEQATESTSLYVGKDLATVFFYMLVALIIHAIQOEYVL
DKINRMHFSKTKSKNESGOLSAFYLFSCINGTFILISENYISDPTILWRAYPHNL
MTFOMKFFIQAQWHPAPPELYFQKEDIPQOLVYIGLYLFTAGAYLLNHL
GLVLVLYHFFYLFELHISRLYFSDERYQKGLWVLFVLGRLLTILSLVLTGFG
ARAENKLDISAGNENLVIAVIAISICITQAFMMKFINFOLRRRREHSTEQAPV
KKKPTVTKGRSSRGTEGVNGVITSGADSPRNKESKSS"
BASE COUNT 830 a 517 c 532 g 977 t
ORIGIN

Query Match 42.0%; Score 961.8; DB 4; Length 2856;
Best Local Similarity 73.2%; Pred. No. 1.1e-221;
Matches 1447; Conservative 0; Mismatches 462; Indels 68; Gaps 14;

Qy 350 gcgacccgggactgactcggatgggaagtggagcccgagctgctacgtgagcgc 409
Db 92 GCGGCACGGCGGGGTGAGAGCAGCCAGCGGGAGCGCGCGCGCGCGCGCAGC 151
Qy 410 ggcgcgtgagagcagccaggggagcgagctgcggctcgcgggtgagtcacgggaag 469
Db 152 CAGTCTGTGAGCAGCTGGAGAGCGCGCGCGGACGAGCACCCTGAGCAGCGCGCGGG 211
Qy 470 cgccacatggggtccgtaagagagacacaaaccccccccgcttctcagccaggaatt 529
Db 212 CCTGCCATTCGCGATTTCGCAAGAGAGCACCAGAGCCCCCGCGTCTCTGAGCCACGAAT 271
Qy 530 catcctgcagaatcatcgggacatcgctcctcgtcggtgggaatgttctcctcgtgggct 589
Db 272 CATCTGCAGAAATCACGGGACATCGTCTCTGTGTGGCGATGCTCTCTCTGCTGGGGCT 331
Qy 590 tgtgttcagggaaacagcagaagcatccatcgtgttctcactctcactctcagcacagtgtgc 649
Db 332 CATGTTCGAGATAACAGCGAAAGCCTCTATCATTTTGTGTACTCTCTCAGTACAATGTTAC 391
Qy 650 tgtccctgcagcagagaaacagccagcggtcctcctcctcctcctcctcctcctcctcctc 709
Db 392 CCTCCCTGCACAGAGAAACAGCTACTGAATCAACATCCCTTTATTACTATGATATCAA 451
```

```
Qy 710 agatttggccacgggttttctctacatgctggtggcaatcattatttcattcgcacaaattca 769
Db 452 AGATTGGCTACAGTTCCTTCTTACATGCTAGTGGCGATAATATTATTCATGCCAATAATCA 511
Qy 770 ggaatatgtgttgataaaataaacaagagatgcagttccaccaagcgcaacaacaaca 829
Db 512 AGAGTATGTGTGGATAAAATAACACAGCGAATGCATTTCTCCAAAACGAAACACAGCAA 571
Qy 830 gtttaacagatggtgcagtttagtgtgtctctacttttttctgttatttggggcacatt 889
Db 572 GTTTAATGAATCTGCTCAGCTTAGTGCATTTACCTTTTCTTTTCTTTTCTTTTCTTTTCT 631
Qy 890 cattttaactctgaaaactccctgcagaccacactcttatatgaaggtcgtcccca 949
Db 632 CATCTAATAATCTGAAAACACTACATCTCAGACCCCAACTATCTGTGGAGGGGTATGCCCA 691
Qy 950 tagcatgatgacatttcaaatgaagtgtttctacatatcccgagttgggttactggtttca 1009
Db 692 TAACCTGATGACATTTCAATGAAGTTTTTCTACATAGCCAGTTGGCTTACTGGTTCCA 751
Qy 1010 tgccttctcgaactctacttccgaaaaacacaaaaacacacacacacacacacacacacac 1069
Db 752 TGCTTTTCTGAACTCTACTTTCCAGAAAACCAAAAAGAGATATTCCTCCTCAGCTTGT 811
Qy 1070 ctacattggtctccactcttccacattcagcattcagcattcctctgttacttgaaatcatt 1129
Db 812 CTACATTGGTCTTTTACCTCTTTACATTCGCGGGCTTATCTTTTAAACTTGAATCACT 871
Qy 1130 gggacttctcttcttggatgcattattttgttgaattacttctccacacatgctcggtct 1189
Db 872 AGGACTTGTCTCTCTGCTGCTACATTTATTTTGTGTAATTTCTCTTCCATATTTCCCGCT 931
Qy 1190 gttttactttagtcataaagaagaccagaaagcatatctctgtggccattgtgtttat 1249
Db 932 GTTTATTTTAGTGATGAAAAGATATCAAAAAGGGTTTCTCTGTGGCGAGTCTGTGTGT 991
Qy 1250 cttgggtagactgtgactttaattgtttccgtactcactgtgtgggtttccacctggctgg 1309
Db 992 TTTGGGAAGACTTCTGACTTTAATCTTTCAGTACTCACGTGTGGCTTTGGCTTTGCAAG 1051
Qy 1310 atcgacagaatcggaatcctgatgccttactggaatgtataatgtgtgtgagcgtataat 1369
Db 1052 AGCAGAGAATCAGAAACTGGATTTTCAATGTGTAGCTGTAGCTGTGTAAGAAAT 1111
Qy 1370 tgcgttctgctgcctccagttgcacgataccagcctcgttaacatggaaacttaactct 1429
Db 1112 TGCTGTCTTGCATCCATTTGCATTTACCAAGCATTCATGATGTGGAGTTCATTAATTT 1171
Qy 1430 ctggctcagagtggtgtagaagattcctaatttcagccctcatgtatgaaaagaacg 1489
Db 1172 TCAGCTTCGAGGTGGAGGGAACATTTACTTTTTCAGGCACCATGTTGTAAGAAAGAAC 1231
Qy 1490 gtc-----gagatcttctaaaaaagacagaaaaacggagt---gggagtgg 1534
Db 1232 AACAGTAACTAAGGCGAGGTCTTCTAGAAAAGGAACAGAAAATGGTGTAAATGGAAACAGT 1291
Qy 1535 aacttcaaatagatgactgctccgcaagagagaaagaaacattctcataa--ctttgc 1593
Db 1292 AACTTCAAAATGGAGCTGACTCTCCCGCTAATAGAAAAGAGAAATCTCATTAATGAATAT 1351
Qy 1594 aagcgattgattaaatgtctgcaaaagaaatcgtctctttgaggtttcttctcgaactaga 1653
Db 1352 AATCTAATTAATGTGCCCCAAAGAAATCTGCTTTTACTATATATCTTTTTCAGCACTAGA 1411
Qy 1654 gatttt---cctgttttgaataatgctcgtcctcctcgttttttttattgaactgttt 1710
Db 1412 GATTTTTCGTCTCTCAAAATACAGTTGTGTCTCTTTGATTCCTTGTATGTACTGTGTTG 1471
Qy 1711 cat-----gtatttttaagacatttgaggggagagattattatgaatgggaaaaaa 1765
Db 1472 TTTTCATGCATTTTTTTAAAGGGCATTTGAGGAAGGATGATTACTATGA--GTGAGAAA 1529
Qy 1766 agattttggtgagactaaattactcatcgtcaaaaataatgtcaaaatagttttgggat 1825
```

```

1530 ATATTTTAGCTAGACTAGCTACCTGCTTCAAAATA-----GTTTAGGAC 1577
1826 caccataattttttgatttttaacaccttcaacatttccctaattgattgacaga 1885
1578 CACCACCAATATCTATTTTGTGTTTT-ATCTTGAATGTTTTCTTAATGATTTGGAGAGA 1636
1886 taact-----gcacaattttgcatcaatgatactggttcttaactccaccagtgcttc 1940
1637 TAACATATTTACAAAATTTCTACATACACAGTATAAAATTTCTGCTCTCACCAATTTTT 1696
1941 ataatactacaagatggtctctctcctcagcaagattatggttttaattgcttgggggt 2000
1697 GTATTAG---CAAGATGGCTCTTCCAGCAAGGTCATTTTTTAAAGTTATCTTTTCAGGCT 1753
2001 aaaaataaagatcagaaaagggtgaagtcaaa-----tcagattctgtaat 2047
1754 AAATGGAATATCTATACAGTTGGAAGTCAAACTTTAAATATGTTTTCAGTGTCTCTAAT 1813
2048 tgtagaattttttttaagaacttacactcagaaaagattgctagactcaccacaaat 2107
1814 TTTTAGGAATT-TTGTAGTCTTTACACCTGGGAAAGAGATTGTAATTCACATATAAT 1872
2108 aataaagtcttatttttacaggtagtgattattagtgcttcoactcccaatttaaaaa- 2166
1873 AAT-TGTGTGCTTTATTTTATAGTAATGTTGTTAATGTTATCATCCCATTTTAAAAAA 1931
2167 ----acacagtaactaatggttaacacatagatgaggttgcgtccatataattgcatca 2221
1932 CAACCAACACATACTATTTGTTTACAAATGTTGGAGAGTTATTGCCATATATATTGAATCA 1991
2222 aaatatcatatttaataataaaattataaatactcctgctccattccactgttaa 2278
1992 AAAATATATGAACATAAAAGTACCTGTGAAGTAAATCTCTTCTTGAGCATTTTCAA 2048

RESULT 5
BC000687 2722 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, translocating chain-associating membrane protein,
DEFINITION clone MGC:784 IMAGE:3347823, mRNA, complete cds.
ACCESSION BC000687
VERSION BC000687.1 GI:12653796
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2722)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.llnl.gov>
 Series: IRAL Plate: 5 Row: 1 Column: 10
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7657654.

FEATURES

source

Location/Qualifiers

1..2722

/organism="Homo sapiens"

/db_xref="LocusID:23471"

/db_xref="taxon:9606"

/clone="MGC:784 IMAGE:3347823"

/tissue_type="kidney, renal cell adenocarcinoma"

/clone_lib="NIH_MGC_14"

/lab_host="DH10B-R"

/note="Vector: pOTB7"

92..1216

/codon_start=1

/product="translocating chain-associating membrane

protein"

/protein_id="AAH00687.1"

/db_xref="GI:12653797"

/translation="MAIRKSKSPVLSHEFVLQNHADIIVSCVAMVFLGLMFEITA

RASLIIFVTLQYNVTPATEEQATESVLSYIGTKDLATVFFVNLVAIIHVAIQEYML

DKINRMHFSKTHSKFSGQSAFYACVWGTFILISENYISDPTILMAYPHNL

MTFQMKFFYISQALYHLAFPELYPKYKKEDIPROLYVIGLYLPHIAGAYLNLNL

GIVLLVHFVEFLFHSRLFYFSNEKYKGFSLMAYLVFLGRLLTLISLVTVGFL

ARAENQKLDFTGNENVLAVRIALASICVTOAFEMMKKEINQLRWREHSFAQAPV

KKKPIVTKGRSSKGTENGVTGLTSLNADSPRNKRKSS"

BASE COUNT 834 a 484 c 480 g 924 t

ORIGIN

Query Match 41.0%; Score 937.6; DB 9; Length 2722;

Best Local Similarity 75.5%; Pred. No. 7.5e-216;

Matches 1394; Conservative 0; Mismatches 389; Indels 63; Gaps 16;

QY 465 ggaagcgccacatcggggtccgtaagaagagcacaagaacccccccgtcttcagccag 524

DB 80 GCGGGCCCCACCATGCGGATTCGCAAGAAAGACCAAGACCCCGGCTGAGCCAC 139

QY 525 gaattcactctgcagaaatcatcgggacatcgctcctcggtgggagatgttcttcctgctg 584

DB 140 GAATTCGCTCTGCAGAAATCAGCGGACATCGTCTCTGTGTGGCGATGCTTCTCTGCTG 199

QY 585 gggctgtgttcgaggaacagcagagatccatcggtgttctcactcttcagcacagt 644

DB 200 GGCTCATGTTTGAGATAACGGCAAGAGCTTCTATCATTTTGTACTCTTCAAGTACAAT 259

QY 645 gttgctgtccctgcagcagaggaacagcagcggtctcaaatccctctattatttgggt 704

DB 260 GTACCCCTCCGACCAACAGAAAGCAAGCTACTGATACAGTGTGCTTATTTACTATGGC 319

QY 705 gtcaagatttgccacggtttctctcactgctggtggaatcattatccatgccaca 764

DB 320 ATCAAGAGATTGGCTACTGTTTTCTTCTACATGCTAGTGGCGATAATTTATTCATGCGTA 379

QY 765 attcaggaatatgtgttgataaaaaataaacaagagagatgcaggttcacaaagcgaaacaa 824

DB 380 ATTCAAGAGTATATGTTGGATAAAATTAACAGGCGAATGCACATCTCCAAAACAAACAC 439

QY 825 aacaagttaacagagctgtgcagtttagtgttctactcttttttttctgtattggtgggc 884

DB 440 AGCAAGTTTAAATGAATCTGTGCTAGCTTAGTGGCTTCTACCTTTTGGCTGTGTTGGGGC 499

QY 885 acattcatttaattctctgaaactcctgctcagaccccaactcttatatggaaggctcgt 944

DB 500 ACATTCTATCTCATCTCTGAARAACATACATCTCAGACCCCAACTACTTATGAGGGCTTAT 559

QY 945 ccccatagcatgatcatttcaaatgaagttttttttatacatatcccatcgttgggtactggt 1004

DB 560 CCCCATAAAGCTGATGACATTTTCAAAATGAAGTTTTTCTACATATACAGAGCTGGCTTACTGG 619


```
Qy 1005 ttccatgcttttccctgaacttacttccagaaaaacacaaacagacatccctcgtcaa 1064
Db 620 CTTCATGCTTTCTCTGAACACTTACTTCCAGAAAACCAAAAGAGATATTCTCGTCAG 679
Qy 1085 cttgtctacattggtcttccactctccacattactcagagcttatctcttgaattgaat 1124
Db 680 CTGTGCTACATGGCTTACTCTTCCACATATGCTGGAGCTTACCTTTTGAACCTGAAT 739
Qy 1125 catttgggaactctctcttggtagctgattatttggtaattacttctccacatgagc 1184
Db 740 CATCTAGACATGTCCTCTGGTGCTACATATTATTTGTGAATTTCTTCCACATTTCC 799
Qy 1185 ggcctgttttacttttagtgatgaagaagaccagaaagccatctctctggtggccattgtg 1244
Db 800 CGCCTGTTTTATTTAGCAATGAAAGATATCAGAAAGGATTTTCTCTGGGCAGTCTCT 859
Qy 1245 tttactctggtagaactgtgaactttaattgtttccgctactcactcgttgggtttccactg 1304
Db 860 TTTGTTTTGGGAAGACTTCTGACTTTAATCTTTTCAGTACTGACTGTGTGGTTTGGCCTT 919
Qy 1305 gctgatacgcagaatcggaatcctgagcccttactcagaaatgtaaatgtgttgagcagt 1364
Db 920 GCAGAGCAGAAATCAGAACTGGATTTTCAGTACTGGAATCTCAATGTGTGTAGCTGTT 979
Qy 1365 aaaattgctgtctgctccagttgcacgacatccaaagcctacgtaacatgaaacttaatt 1424
Db 980 AGAATCGCTGTTCTGGCATCCATTTGCGTTACTCAGGCGATTTATGATGGAACTTCATT 1039
Qy 1425 actctcgtgcttcagagtggtgagaagattcttaataattccaggcctcactgataaaaaag 1484
Db 1040 AATTTCAGCTTCGAAGGTGGAGGGAACATTTCTGCTTTTCAGGACCACAGCTGTGAAGAAG 1099
Qy 1485 aaacggctcg-----agatcttctaaaaaagaacagaaacggagtg---gga 1529
Db 1100 AAACCAACAGTAACTAAGGCGAGATCTCTAAAAAAGGAACAGAAATGGTGTGAATGGA 1159
Qy 1530 gtggaacctcaaatagatgtagtgcgcgaagaggaagaaagaaatcttcataat-c 1588
Db 1160 ACATTAATCTCAAAATGTAGCAGACTCTCCCGGAATAAAAAAGAGAAATCTTCATAATGA 1219
Qy 1589 ttgcaagcgcattgattaatgtctgcaagaagaaatcgtctcttggaggttcttcttgca 1648
Db 1220 AATTAAACTAATTGATTAATGTCCCAAAAGAAATCTGCTTCTACTATATCTTTTCAGCA 1279
Qy 1649 ctagagatttttctgtttttgaaaaat--agttcgtgctcttcggttttggttattgaaact 1706
Db 1280 TTAGAGATTTTCTGCTTCTGAAATACAGTCTGCTCTTTGATTTTGGTATGTAGC 1339
Qy 1707 gttcctgattttttt--aaagacatttgaggaggaggaattattatgaatgggaaaaa 1764
Db 1340 GTTTCATGCATTTTTTTAAAGGGCATTTGAGGGGAGGATTTTGTCTATGAAT--GAAAAA 1397
Qy 1765 aagatttgggtgagactaaattactcactcgtcaaaaataatgcaaaaatggtttgggga 1824
Db 1398 AATATTTTAGCTTAGACTAAGCTACCTGCCCTCAAAATA-----GTTTAGGGA 1445
Qy 1825 tcaccactatatttggttgatttttaacaccttcaacattttcctaattgatttgcagag 1884
Db 1446 CCACCACCATATTATTTTTCG--TTTTTATTTTTCACATTTTTTCTAATGATTTGGAGAG 1503
Qy 1885 ataactgc-----acaatttgcatacaatgatactcgtgttcttactccccaccagtggtt 1939
Db 1504 AAAAATAATTTACAAAAATTCACATATACAGTGATACAAATTTCTTGTCTCACCAATTTT 1563
Qy 1940 cataactaacaagatggtctcctcctcagcaagattatggttttaagtctgtctttt-ggg 1998
Db 1564 TATAA---TAGCAGAGTGGCCTGTCTTAAGAAAGGCCATATTTTTTAAAGTATATCTTTCCAGG 1620
Qy 1999 gtaaaaataaaagtcagaaaaagtggaagtcaaa-----tcagtatcttgta 2045
Db 1621 GTACATGGAATACTATAAAGTTGGATGTCAAACTTAATATGTTTTTCAGTGTCTCTA 1680
Qy 2046 atgttagaattttttttaagaacttacaactcagaaaaagattgttagactcaccacaa 2105
```

```
Db 1681 ATTTTTCGGAATT-TTGTAGACTTTTACACCTGGAAAAAAGATTGTGAAATCACCAGA 1739
Qy 2106 ataataaagtctcttatttttacaggtagtagtattattagcttctcattt--aaa 2163
Db 1740 ACAAT-TGTGTGCTTTATTTTATAGGTAGTGGTTATTAGTATTACATCCCATTTTAAAA 1798
Qy 2164 aaacacagctactaatgggtaacacatagaggttttgcctccatataattgcataaa 2223
Db 1799 ACAAAAACATTAATGTTTACACACCTGGAGTTTACTTAACATACATATTAATATCAA 1858
Qy 2224 atatcat-taataataaaaaattattaaaaatcattctctctccatt 2268
Db 1859 GTATATTCTTAAAAGTACTTCTGGAAGTAAATCTTCTTGTGCATT 1904

RESULT 6
BTU19578 2387 bp mRNA linear MAM 06-SEP-2000
LOCUS Bos taurus translocation chain-membrane associating protein (tram)
DEFINITION mRNA, partial cds.
ACCESSION U19578
VERSION U19578.1 GI:9971727
KEYWORDS
SOURCE
ORGANISM Bos taurus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Clark,T.G., Morris,J., Akamatsu,M., McGraw,R.A. and Ivarie,R.D.
TITLE Cloning and sequence analysis of a bovine tram cDNA
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 2387)
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1995) Genetics, University of Georgia, Athens, GA
30602-7223, USA
FEATURES
source 1..2387
/organism="Bos taurus"
/db_xref="taxon:9913"
<1..2387
/gene="tram"
<1..1077
/gene="tram"
/codon_start=1
/product="translocation chain-membrane associating
protein"
/protein_id="AAG10391.1"
/db_xref="GI:9971728"
/translation="EFVLONHADIVSCVMVFLGLMFEITAKVSIIFVTLOYNTLP
ATPEQATESAFLYYGIKDLATVFFYMLVAIIIIHAIQEVYLDKINRMHFSKTKHSK
FNESGQLSAFLFSCWIFLISENYITLWRAYPHNLMTFQMKFYISOLAYH
FHAFPELYFOYKKEIDIPQOLVYGLYFHAGYLLNLHLGLVLLVLYFVFLPH
ISRLFFSDEKYQKQSLWAVLVGLRLLLILSLVTVGFGLAENOKLDFSTGFN
VLAVRIATLASICTOAFMMKFTINFLRRWRHSAFQAPAVKKKPPVTKGRSXXKGT
ENGVNCITVTSNGADSPNRKESK"
BASE COUNT 675 a 473 c 449 g 788 t 2 others
ORIGIN
Query Match 35.2% Score 804.8; DB 4; Length 2387;
Best Local Similarity 76.6% Pred. No. 9.7e-184;
Matches 1137; Conservative 0; Mismatches 309; Indels 38; Gaps 11;
Qy 525 gaattcatctgcagaatcatcgagacatcgtctcctcgtcgtgggagatgttctctcgtcgt 584
Db 1 GAATTCGCTCGCAGATCAGCGGACATCGTCTCCTGCTG7GCGATGGTCTTCTCTGCTG 60
Qy 585 gggcttggttcgagggaacagcagacatcccatcgtgttctcactcttcacagacagt 644
Db 61 GGGTCATGTTTCGAGATAACGAAAAAGTTTCTCATCATTTTTTGTACTCTTTCAGTACAAT 120
```


QY 645 gttgctgtccctgcagcagaggaacacagccaggggtccaaagtccctctattattatggt 704
Db 121 GTTACCCTCCCTGCCACAGAACAGCAAGCTACTGAATCAGCGTCTCTTTATTACTATGCT 180
QY 705 gtcaagaatttgccacaggggtttctctctacatgctggtggcgaatcattatccatgcaca 764
Db 181 ATCAAGATTTGGCTACAGCTTTCTTCTACATGCTAGTGGCGATAATATTATCATGCGATA 240
QY 765 attcaggaatatgtgtgataaaatacaagagaaatcaagtcacccaagcgaacaa 824
Db 241 ATTCAGAAATATGTTGGATTAATAATTAACAGCGGAATGCATTTCTCCAAAAGCAACAC 300
QY 825 aacaagtttaacagagctggtcagtttagtgttctactctttttcttctgtatttggggc 884
Db 301 AGCAAGTTTAAATGAATCTGGTCAGCTTAGTGCATCTACCTGTCTTTTCTGTTTGGGCG 360
QY 885 acattcattttaatctctgaaacacgctgtcagaccccaactcttattatggaagcctgt 944
Db 361 ACATTCAATCTCATATCTGAAACACTACATCTCAGACCCGACTATTTTGTGGAGGGCTTAT 420
QY 945 ccccatagcatgatacaatttcaaatgaagtttttctacatccatccagttggtctactgg 1004
Db 421 CCCATTAACCTGATGACATTTTCAGATGAAGTTTCTTACATATCAGATTTGGCATACTGG 480
QY 1005 ttctcatgcttttctgaaactctacttccagaaaccccaaaacagacacatccctcgtcaa 1064
Db 481 TTTTCATGCTTTTCTGAACTCTACTTCCAGAAACCAAAAAGAGATATTCTCTGCTCAG 540
QY 1065 cttgtctacatggttcttccactcttccacattactgagagcttatctctgttactgaat 1124
Db 541 CTCGCTCATATCGGCTTTTACCTCTTTCATTTCTGAGCTTACCTTTTAAACTTGAAT 600
QY 1125 catttgggaacttctcttggtaactgcaattatttttggtaattactttccacatgtgc 1184
Db 601 CATCTAGGACTGTGCTTTGGTGTCTACACTATTTTGTGAGTATTTCTTTTCCACATTTCC 660
QY 1185 ggcctgttttacttagtgataaagaccagaaagggacatactctctgtgggccaattgtg 1244
Db 661 CGCCTATTTTACTTAGTCACGAAAGATATCAGAAAGGATTTTCTGTGGCAGTTCG 720
QY 1245 ttattctgggtagactgtgactttaattgtttccgtactcaactgttgggtttcacctg 1304
Db 721 TTTGTGTTGGGAAGACTCTGACTTAAATCTCTCGGTACTCAGTCTGGCTTTGGCTTA 780
QY 1305 gctgtagcgaatcgaaactcctgactcccttactgaaatgtaaatgttggcagct 1364
Db 781 GCAAGAGCAGAACGACGAGGCTGGATTTTCAGTACTGGAACTTCAATGTGTAGCTGTT 840
QY 1365 aaaaattgctgttctgtcgtccagttgcacgataccagcctacgttaacatgggaacttaatt 1424
Db 841 AGAATTGCTGTTCTGGCATCCATTTGCATTTACCAAGCATTTATGATGTGGAAGTTCATT 900
QY 1425 actctctggctcagagtggtgtagaagattcattatccagcctcattgtatgaaaaag 1484
Db 901 AATTTTACGCTCGAAGTGTGAGAGAACATTTCTGCTTTTTCAGGCCACCGCTGTGAAGAAG 960
QY 1485 aaacggtc-----gagatcttctaaagaaagaaacgaacgagct---ggga 1529
Db 961 AGGCCACCATGCTACTTAAGGCGAGGCTCTCTAGNAAAGGAAACAGAAATGGTCTCAATGGA 1020
QY 1530 gtggaaacttcaaatagatgactgtcgcacgaagagaaagaaagaaacttctcataat-c 1588
Db 1021 ACAGTAATCTCAAAATGGACGACACTCTCCCGCTAATAGAAAAGAGAAATCTTCATAATGA 1080
QY 1589 ttgtcaagcgcattgatttaattgtcgaaggaattctgctctcttgggtttcttctgca 1648
Db 1081 ATTATAATCTAATTAATGTAATGTCCCAAGAAATCTGCTTTTACTATATCTTTTCAGCA 1140
QY 1649 ctgagatttttctgtttt---tgaaaaatagttgctctcttgggtttttg---ttatt 1701
Db 1141 CTAGAGATTTGTTTCTGTTCTTCTGAAATACATTTTCTGCTCTTGTATTCTTCTACTATT 1200

QY 1702 gaactgttttcattgatttttttaaa--gacatttgaggaggagattattatgaatggg 1759
Db 1201 GTATTATTTTCATGCAATTTTAAAGGGCATTTGAGGGAGGATGATTCCTATGA--GTG 1258
QY 1760 aaaaaaagatttgggtggagactaaattactcatcgtcaaaaataatgcaaaaatagtttt 1819
Db 1259 AAAAAAATGTTTGTAGTCTTAAGCTACCTGCTTCAAAATAGTGTAGGGACCAACAT 1318
QY 1820 gggggtaccacactatatttggtttggatttttaacctttcaacattttcttaagtattg 1879
Db 1319 ---CATATTATATTTTATTTTATTTTGTGTTTTTATCTTTAAGCATTTTCTAATGATTG 1375
QY 1880 cagagataact-----gcacaatttgcatacaatgat-actgggttcttactccacca 1933
Db 1376 AAGAGATAACTATTTTACAAAATTTATACATATCAGTGATAAACAATTTCTTGTCTCACCA 1435
QY 1934 gtgtttcataataactaacaagatggtctctcctctagacaagattat 1977
Db 1436 ATTTTATATTA--AGCAAGATGGCTGTTCAGCAGGGGTCAAT 1477
RESULT 7
AC106086
LOCUS
DEFINITION
AC106086 155623 bp DNA linear HTG 12-JAN-2002
Rattus norvegicus clone CH230-126G18, *** SEQUENCING IN PROGRESS
***, 70 unordered pieces.
ACCESSION
AC106086
VERSION
HTG: HTGS-PHASE1.
KEYWORDS
Norway rat.
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 155623)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbara, J.,
Benton, J., Blmage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, I., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newsham, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogih, M., Okunou, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Roostari, N.,
Rulz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoofat, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalobos, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williams, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 155623)
Worley, K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHHR
Center clone name: CH230-126G18
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 127470 bases at least Q40
Consensus quality: 134419 bases at least Q30
Consensus quality: 139475 bases at least Q20
Estimated insert size: 122947; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 70 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 5824: contig of 5824 bp in length
5825 5924: gap of unknown length
5925 10293: contig of 4369 bp in length
10294 10393: gap of unknown length
10394 13928: contig of 3535 bp in length
13929 14028: gap of unknown length
14029 19439: contig of 5410 bp in length
19439 19538: gap of unknown length
19539 22725: contig of 3187 bp in length
22726 22825: gap of unknown length
22826 24943: contig of 2118 bp in length
24944 25043: gap of unknown length
25044 29216: contig of 4173 bp in length
29217 29316: gap of unknown length
29317 33015: contig of 3699 bp in length
33016 33115: gap of unknown length
33116 37717: contig of 4602 bp in length
37718 37817: gap of unknown length
37818 40275: contig of 2458 bp in length
40276 40375: gap of unknown length
40376 42400: contig of 2025 bp in length
42401 42500: gap of unknown length
42501 45023: contig of 2523 bp in length
45024 45123: gap of unknown length
45124 48116: contig of 2993 bp in length
48117 48216: gap of unknown length
48217 51729: contig of 3513 bp in length
51730 51823: gap of unknown length
51830 54589: contig of 2760 bp in length
54590 54689: gap of unknown length
54690 57532: contig of 2843 bp in length
57533 57632: gap of unknown length
57633 60697: contig of 3065 bp in length
60698 62618: gap of unknown length
62619 62718: gap of unknown length

62719 65151: contig of 2433 bp in length
65152 65251: gap of unknown length
65252 67843: contig of 2592 bp in length
67844 67943: gap of unknown length
67944 69264: contig of 1321 bp in length
69265 69364: gap of unknown length
69365 71847: contig of 2483 bp in length
71848 71948: gap of unknown length
71949 74812: contig of 2865 bp in length
74813 74912: gap of unknown length
74913 77983: contig of 3071 bp in length
77984 78083: gap of unknown length
78084 80155: contig of 2072 bp in length
80156 80255: gap of unknown length
80256 81813: contig of 1558 bp in length
81814 81913: gap of unknown length
81914 83742: contig of 1829 bp in length
83743 83842: gap of unknown length
83843 86103: contig of 2261 bp in length
86104 86203: gap of unknown length
86204 88411: contig of 2208 bp in length
88412 88511: gap of unknown length
88512 90397: contig of 1886 bp in length
90398 90497: gap of unknown length
90498 92759: contig of 2262 bp in length
92760 92859: gap of unknown length
92860 94270: contig of 1411 bp in length
94271 94370: gap of unknown length
94371 96506: contig of 2136 bp in length
96507 96606: gap of unknown length
96607 97996: contig of 1390 bp in length
97997 98096: gap of unknown length
98097 99954: contig of 1858 bp in length
99955 100054: gap of unknown length
100055 101350: contig of 1296 bp in length
101351 101450: gap of unknown length
101451 103336: contig of 1886 bp in length
103337 103436: gap of unknown length
103437 104996: contig of 1560 bp in length
104997 105096: gap of unknown length
105097 107126: contig of 2030 bp in length
107127 107226: gap of unknown length
107227 108668: contig of 1442 bp in length
108669 108768: gap of unknown length
108769 110348: contig of 1580 bp in length
110349 110448: gap of unknown length
110449 112318: contig of 1870 bp in length
112319 112418: gap of unknown length
112419 114291: contig of 1873 bp in length
114292 114391: gap of unknown length
114392 116376: contig of 1985 bp in length
116377 116476: gap of unknown length
116477 118335: contig of 1859 bp in length
118336 118435: gap of unknown length
118436 119947: contig of 1512 bp in length
119948 120047: gap of unknown length
120048 121943: contig of 1896 bp in length
121944 122043: gap of unknown length
122044 123766: contig of 1723 bp in length
123767 123866: gap of unknown length
123867 125034: contig of 1168 bp in length
125035 125134: gap of unknown length
125135 126827: contig of 1693 bp in length
126828 126927: gap of unknown length
126928 128578: contig of 1651 bp in length
128579 128678: gap of unknown length
128679 129702: contig of 1024 bp in length
129703 129802: gap of unknown length
129803 132175: contig of 2373 bp in length
132176 132275: gap of unknown length
132276 134086: contig of 1811 bp in length
134087 134186: gap of unknown length
134187 135264: contig of 1078 bp in length

	* 135265	135364:	gap of unknown length	
	*	135365	136997: contig of 1533 bp in length	
	*	136898	136997: gap of unknown length	
	*	136998	138452: contig of 1455 bp in length	
Query Match	35.1%	Score 803.8;	DB 2;	Length 155623;
Best Local Similarity	68.3%;	Pred. No. 4.9e-183;		
Matches 1419;	Conservative	0; Mismatches 562;	Indels 98;	Gaps
QY	167	cgtgacgcagcagatttgaagtatcgacatacaaoqttagaatcacgcctcgcaaccocq	226	
Db	1592	CGGCTCCGCGGTGGCGCGCGCATCGGTGCGCAAAATCCGGCTCGCACCGCGG	1651	
QY	227	ggcctggcgccgggtaggggcgcgctgatcttccttccttcgcctccgcctccccct	286	
Db	1652	-----CCTGCGGGGACGGCGCGCGTCTGATTTCCTTCCTGCCCTCTGCAG---	1703	-CCCC
QY	287	ggtgcgatgtccagctcaagctcgcccctcgcccttgatttttttttttttttttttt	346	
Db	1704	TGTGCCATGCTCGGC-CFACCGCGCCCAGCCTTTGATTGA-----TCGGTCGGCAGCG	1757	
QY	347	gctgcgacccggagactgcctcggaatgggaagtggagcccccggagctgctacnctg	406	
Db	1758	GCTCGACACTGGCGGCAGACGGCGGGATGGGASCC-----CGGCGCTGGG	1807	
QY	407	ggcgcgctgtgagagagccaggggagcagctcggtcgcccggtgagtatccogg	466	
Db	1808	AGCGCGCGAGTGATCAGC-----GGTGGCGCGGTGAGTAGTACCGGTGAACCGCGG	1859	
QY	467	aegcgcccatggggctcctaagaagagaccaaagaacccccccgttctcagcagga	526	
Db	1860	-----CATTGGGGCTCCGCAAGAAGAACCCAGGAACCCCGGTGTGAGCCACGA	1910	
QY	527	attcctctgcagaatcatgcggacatogtctctcggtgggagtggtctctctgctgg	586	
Db	1911	ATTATGTTGTCAGAACACCGGATATGTTCTCTGCTGGCATGTTCTTCTGCTGGG	1970	
QY	587	gcttgtttcagggaaacagaaagatccatcgtgtttctcactcttcagcaacagt	646	
Db	1971	ACTTATGTTTCGAGGGCACGGCGAGATGTCATCGTGTTCCTACCTCGCAGCATGGAGT	2030	
QY	647	tgctgtccctgcagcagaggaacaagccagcggtccaagtccctctattattatggt	706	
Db	2031	CGTTGTCC-- -CAGCGAAGGGCTACCTCGGGTCCAGGACCCTTACCATATGGGT	2087	
QY	707	caagatttggccacggttttcttcacatgctggtggcaatcattattccacacat	766	
Db	2088	CAAAGATCTGGCCACAGTGTCTTCTACATGCTGTGGCCATCATATTACGCCACCAT	2147	
QY	767	tcaggaaatgtgtggataaaaattaacaagagaatcagttcaccaagcgaacaaaa	826	
Db	2148	TCAGGAGTACGTGTAGATAACTCAGCGGAGACTGCGAGCTACCAAGGCAAAACAAA	2207	
QY	827	caagtttaacgagctgtggttcagtttagtgttctactttttttctgtatttggggcac	886	
Db	2208	CAAAATTGAATAGGCGGSCACTCAGTGTGTTTACATAGTGTCTGTATCTGGGTAT	2267	
QY	887	attcattttaactctgnaaacctgctdgcagcccaactcttatatggaggtcgtcc	946	
Db	2268	GATCATCTGGGCTCTGAGAACTGCTGTACAGCCCACTCTATTGTGGAGTCTCAGCC	2327	
QY	947	ccaatagatgatcaatttcaaatgaatttttacaatatccagttgacttacgtgtt	1006	
Db	2328	CCACAACATGATGACATTTTCAGATGAATTTTTTCTACATCTFCAGTGTGCTTACTGGTT	2387	
QY	1007	tcatgtctttctgaactctacttccagaaaaaccaaaaaaagaacatccctcgtcaact	1066	
Db	2388	TCATAGTTTCCCGAGGCTACTTCCAGAAAGTCAGGAAAAAAGATATCCGGGTCAACT	2447	
QY	1067	tgtctacatgtgcttcaacctctccacattacttgagcttatctcttgtaactaatca	1126	
Db	2448	CATCTACATTGGCCCTCCAGCTTCCACATTGGAGGGCCCTATCTCTGTACTTGAACCA	2507	

Qy	1127	t t g g g a c t t c t t o t t t g g t a c t g a t k a t t t t g t t g a a t t a c t t t c c a c a t g t o g g	1186
Db	2508	C C T G G C G T G T G T T C T G A T G T G C A C T A T G C T G C A G C T C C T C C A G C G T G T G C A G	2567
Qy	1187	c g t g t t t a c t t i a g t g a t g a a a g t a c c a g a a a g g c a t a t c t g t g g c c a t t g t t	1246
Db	2568	C C T G C T T T A C T T T G G G G A T G A G C G G T A C C A G A A A G G T T G C T T T G T G G C C T A T C G T G T	2627
Qy	1247	t a t c t t g g g t l a g a c t t g t g a c t t t a a t t g t t t c c g t a o c t a c t g t t g g g t t c a c c t g g c	1306
Db	2628	T A T A T C C G G G A G A C T C G T G A C A C T G A T T G T C T A G T G G T T A C A G T A G G C T T C A C T T G G C	2687
Qy	1307	t g g a t c g a a t o g g a a t c o t g a t g c c e t t a c t t a c t g g a a a t g t a a a t g t t g t t g g c a g t a a	1366
Db	2688	C G G - - - G A C A A A T C G G A A T G G A A A T G C T C T C T G T A A T G T C A A T G T G T T G G C A G C T A A	2744
Qy	1367	a a t t g t g t t c t g t c c a g t t g e a g a t c c a a g c c a c g t a a c a t a a a t g a a a c t t a a t t a c	1426
Db	2745	A A T C G C T G T T C T G C T C G A G T T C A G A G T T C A G A G T T A C A P A A C A T G G A C C T T G A C A C	2804
Qy	1427	t c t c t g g c t t c a g a g t t g g g t a g a a g a t t c t a a t t c a g g c c t c a t g t a t g a a a a g a a	1486
Db	2805	C G T C T G C C T C A G A T G G T T A G A A G A T C G G A A T C T T C A T G C T C T G G G A G A A - - - - -	2858
Qy	1487	a c g g t c g a g a t c t t c t a a a a a a g a a c a a c g g a g t g g g a g t g g a a a c t t c a a a t a g	1546
Db	2859	- - - - - G A G A C G G T C C A G G T C G A G A A A A G G C A C A G A A A A T G G A A T C C A A A T A G	2912
Qy	1547	a g t a a c t g t c c g c c a a a g a a a g a g a a a t c t t c a t a a t c t t t g c a a g c a t t g a t t	1606
Db	2913	A A T A G A T T C T C C A C C A A A G A A A G A A A A G A A A G C T C C T T A G C A G T T T G C A A G C G A A T T G A T T	2972
Qy	1607	a a t g t c t g c a a a g a a t - - - c t c t c t t t g a g g t t c t t t t g c a c t a g a g a t t t t c t g	1663
Db	2973	C T T A C C T C C A A G G A A T C C A C T T C T T C T A T G T G G T G C T C T G T C C T A G A - T T T T C T G	3031
Qy	1664	t t t t g a a a t a g t t o g t o t c t o g g t t t t g t t a t g a a c t g t t c a t g a t t t t t t a	1723
Db	3032	T T C T T C A G A A C G G T C G T G C T T T T T G A A T A T T G C T A A T G T A T T G T C T A A T G T G T T T T - -	3089
Qy	1724	a a g a c a t t g a g g g g a g g a t t a t t a g a a t g g a a a a a a a a g a t t t t g t t g a g a c t a	1783
Db	3090	- - - - - A A G T T T T C G A C A C G T A T A G T G G G G A T G G G G T T A A G A C T A	3132
Qy	1784	a a t t a c t a c t g t c a a a a a a t g c a a a a t a g t t t t g g g a t c a c c a c t a t a t t t t g t t	1843
Db	3133	A A C C A C T C A G C C T T A A N A C T G T C A G A T A G T T A A C G G - - - A C C A C A C T T A T T A T T A G	3188
Qy	1844	t g a t t t t a a c c t t c a a c a t t t c c t a a t a t t g t c a g a g a t a a a c t g c a c a a t t t g c a	1903
Db	3189	T T A G G T T C T T A C C T C A A C A G A T T T T C C A A A C G T T T T G T G T A T G A C T G C A G A A T T G T A	3248
Qy	1904	t a t a a t g a t a c t g g t t c t t a c t c c a c c a g t t t t c a a t a a c t a a c a a a g a t g g t c t c t	1963
Db	3249	C A T A A A T A A N A F A G N - - - - T T C C T G C T T C C A A T G T T C T T T A T G C A A T T A A C A A G T C T C T A	3303
Qy	1964	c c t a g c a a g a t a t g t g t t a a t g c t g t g t t g g g g t a a a a t - a a a a g t a c g a a a a a g g	2021
Db	3304	G C A A A G T G G T T G T T T C T C A A T G T T C T C T C C A G G A T A A A G T G G A A A T C T G A T A A A G G	3363
Qy	2022	t g g a a g t c a a a t c a g t a t t c t g t a a t t g t t a g a a t t t t t t t a a g a a c t t a c a a c t a	2081
Db	3364	T T A A A C T C A A A T C A G T A T T A T G T A A C C G T T G G A T T T T T T A A A G T T T T A A A T T T A C A	3423
Qy	2082	g a a a a g a t g c t a g a c t a c c a a a a a t a a t a a a t g t t c t t t a t t t t a c a g g t a g t g a t t a t	2141
Db	3424	A T G A A A A G A C A T T T G T C A A A C C A C C A - A A A A A T A T G T G T T A A T T T T A T G A G T A G T A A T T G T	3482
Qy	2142	t a g t g t c t a c c c a t t t a a a a a a c a c a g t a c t a a t g g t a a c a c a t a t g g a g g t t t g	2201
Db	3483	T A G T G C T T A C G C C C C A T T A A A G C A T C A A A A T A G A A T A G A T G A C A T G T G T G G T G A T - -	3541
Qy	2202	c t g c a t a t a t a t t g c a t c a a a a t a c t a a t t a a t a t	2240

```
Db 3542 ---TGACATTTAGCGAATCAAGATACCTTTAATAAATAT 3577
|||||
RESULT 8
AX329795
LOCUS AX329795
DEFINITION Sequence 304 from Patent WO0194629.
ACCESSION AX329795
VERSION AX329795.1 GI:18102773
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horligan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 304 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Location/Qualifiers
1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN
Query Match 32.4%; Score 742.4; DB 6; Length 1267;
Best Local Similarity 79.7%; Pred. No. 1e-168;
Matches 907; Conservative 0; Mismatches 216; Indels 15; Gaps 2;
QY 465 ggaagcgccaccatggggctccgttaagaagcaccacaaagaccccccgttctcagcag 524
|||||
Db 110 GCGCGCCCCACCATGCGATTGCGAAGAAAGCACCAGAGACCCCGCTGAGCCAC 169
|||||
QY 525 gaattcatctgcagaataatcgagacatgctctcgtcggtgggagtgttctcctgctg 584
|||||
Db 170 GAATTCGTCCTCGAATACAGCGACATCGTCTCTGTGGGATGGTCTTCTCTGCTG 229
|||||
QY 585 gggctgtgttcagggagcaggaagcattccatcggtttctcactcttcaacacagt 644
|||||
Db 230 GGGCTCATGTTTGAGATAACGCGAAAGCTTCTATCAITTTTGTACTCTTCAGTACAAT 289
|||||
QY 645 gtgtgctgcttcgagcagaggaacgcccgggtccaaagtcctctattattatggt 704
|||||
Db 290 GTCAACCTCCAGCAACAGAAAGAACTACTGAATCAGTGTCCCTTTATTACTATGGC 349
|||||
QY 705 gtcaagagatttggcagcggttttctctacatgctgtggaatcattatctatgcaca 764
|||||
Db 350 ATCAAGATTTGGCTACTGTTTCTTCTACATGCTAGTGGCGAATAATATTATGCGCGTA 409
|||||
QY 765 attcaggaataatggttggaataaattacaagagaatgcagttcaccaagcgaaacaa 824
|||||
Db 410 ATTCAAGAGTATATGTTGGATAAAATTAACAGCGGAATGCACATTCACAAACAAACAC 469
|||||
QY 825 aacaagtttaacgagctgctcagtttagtgttcttaacttttttcttctgtattggggc 884
|||||
Db 470 AGCAAGTTTAATGAATCGTGCAGCTTAGTGCCTTCTTACCTTTTGTGCTGTGTTGGGC 529
|||||
QY 885 acattcatattatctctgaaaactgctgtcagaccccaactcttatatgggaagcctct 944
|||||
Db 530 ACATTCATCTCTCTCTGAAACTACATCTCAGACCCCACTATCTATGAGGGCTTAT 589
|||||
QY 945 ccccatgacatgatgacatttcaatgaagtgttttttctacatatccagttggtctactgg 1004
|||||
Db 590 CCCCATACTGATGACATTTCAATGAAGTTTTCTACATATCAGCTGCTTACTGG 649
|||||
QY 1005 ttctatgcttttctgaactctacttccagaaaacacaaacaaagacatccctcgtcaa 1064
|||||
Db 650 CTTTCATGCTTTTCTGAACTCTACTTCCAGAAACCAAAAGAGATATTCCTCTGTCAG 709
|||||
```

```
QY 1065 cttgtctacattggtcttccaccttccacattactgagcgttactctctgttacttgaat 1124
|||||
Db 710 CTTGTCTACATTTGGTCTTTTACCTCTTCACATTCGTGGAGCTTACCTTTTGAACCTTGAAT 769
|||||
QY 1125 catctgggaactctctcttcttggactgacttatttttggtaattacttccacatgtgc 1184
|||||
Db 770 CATCTAGGACTTGTCTTCTGTGCTGCTACATATTATTGTGTAATTTCTTTTCCACATTTCC 829
|||||
QY 1185 ggcctgttttacttttagtgaagaaagaccagaaagcattatctctgtgggacctgtg 1244
|||||
Db 830 CGCTGTTTTATTATTAGCAATGAAAGATATCAGAAAGGATTTTCTCTGTGGCGAGTTCTT 889
|||||
QY 1245 ttatcttgggtagactgtgactttaaattgttccgtactcactggttgggtttcaacctg 1304
|||||
Db 890 TTTGTTTTGGGAAGACTTCTGACTTTAATTTCTTTCAGTACTGACTGTGTGTTTGGGCTT 949
|||||
QY 1305 gctgagtcgagaatcggaatccctgagtcgacctactggaatgtaaaatgtgttggcagct 1364
|||||
Db 950 CCAAGAGCAGAAATCAGAAACTGGATTTTCAGTACTGGAACCTCAATGTGTAGCTGTT 1009
|||||
QY 1365 aaaattgctgtctgtcgtccagttgcacgattccaaagcctacgtaacatggaacttaatt 1424
|||||
Db 1010 AGAATCGCTGTTCTGGCATCCATTTGGTTACTCAGGCATTTATGATGTGAAGTTCAAT 1069
|||||
QY 1425 actctggtctcagaggtggttagaagattcttaattcagggcctcattgatgaaag 1484
|||||
Db 1070 AATTTTCAGCTTCGAAGTGGAGGGAACATCTGCTTTTCAGGCACCCAGCTGTGGAAGAG 1129
|||||
QY 1485 aaacggctc-----gagatcttctaaaaaagaagaacagaaacggagtg---gga 1529
|||||
Db 1130 AACCAACAGTACTAAGCGCATCTCTTAAAGAAAGGAACAGAAATGCTGTGAATGGA 1189
|||||
QY 1530 gtggaacttcaaatgagtagactgtccgcacaaagagaaagaaatcttctcataat 1587
|||||
Db 1190 ACATTAACTTCAATGTAGCAGACTCTCCCGGAATTAATAAAGAGAAATCTTCATAAT 1247
|||||
RESULT 9
BD005229 1267 bp linear PAT 31-JAN-2002
LOCUS Method of screening of protein.
DEFINITION BD005229
ACCESSION BD005229
VERSION BD005229.1 GI:18633190
KEYWORDS JP 03075332-T/4.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1267)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
AUTHORS Method of screening of protein
TITLE Patent: JP 03075332-T 4 16-FEB-2001;
JOURNAL SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA,OTOAKI
IMAMURA,HIROKORI ISHIKAWA,KIYOMITSU NEMOTO
COMMENT OS Homo sapiens (human)
PN JP 03075332-T/4
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PR 20-AUG-1999 JP 99P 234764
PI NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIROKORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C1201/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
(C12P21/02,C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1267
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 343 a 278 c 275 g 371 t
ORIGIN
```

[illegible]

Qy	645	gttgctgtccctgcagcagaggaacagccagcggtccaaagtccctctattattatggt	704
Db	290	GTACCCCTCCAGCAACAGAGAACAGCTACTGAATCAGTGCCTTATTACTATGGC	349
Qy	705	gtcaaaagatttggccagcggttttctctacatgctggtggcaatcatattatcagccaca	764
Db	350	ATCAAGAATTGGCTACTGTTTCTTCTACATGCTAGTGGCGATAAATTATTCATGCCGTA	409
Qy	765	attcaggaatatgtgttgataaaattaaacaagagaatgcagttccacaaaggaacaa	824
Db	410	ATTCAAGAGTATATGTTGGATAAAATTAACAGGCGAATGCATTTCTCCAAAACAAACAC	469
Qy	825	aacaagtttaacagagctggcagcttagtggcttacttttttcttatttttctatttggggc	884
Db	470	AGCAAGTTAATGAATCTGGTCACCTTAGTGGCTTCTACCTTTTGGCTGTGTTGGGGC	529
Qy	885	acattcatttaactctgtaaaactgcctgtcagaccacaaactcttatatggaaggtcgt	944
Db	530	ACATTCAATCTCTCTCTGAAACTACATCTCAGACCCAACTATCTTATGGAGGCTTAT	589
Qy	945	ccccatgacatgacatttcaaatgaagtttttctacatatcccaagttggcttactgg	1004
Db	590	CCCATACCTTGATGACATTTCAATGAAGTTTTCACATATCACAGCTGGCTTACTGG	649
Qy	1005	tttcatgcttttctgaactctacttccagaaacccaaacaaagacatccctcgtaa	1064
Db	650	CTTCATGCTTTTCTGTAACCTACTTCCAGAAAACCAAAAAGAAAGATATCTCTCGTCAG	709
Qy	1065	cttgtctacatgtgtcttcaactctccacattactgtgagcttatctcttgtacttgaaat	1124
Db	710	CTTGTCTACATTTGCTTTACCTCTCCACATTTGCTGGAGCTTACCTTTTGAACCTTGAAT	769
Qy	1125	catttgggaactctcttgggtactcattatttttgaattactttccacatgtgc	1184
Db	770	CATCTAGACCTGTCTCTGGTGTCTACATATTTTGTGAATTTCTTTTCCACATTTCC	829
Qy	1185	ggcctgttttactttagtgaaagtaccagaaagccatatactctgtggggcattgtg	1244
Db	830	CGCCTGTTTATTTAGCAATGAAAGTATCAGAAAGGATTTTCTCTGTGGCAGTTCTT	889
Qy	1245	tttatctgggtagacttgaactttaattgtttccctactcactgctgtggttttcaactg	1304
Db	890	TTTGTGTTGGAGACCTCTGACTTTAATCTTTTCAGTACTGACTGTGTGTTTGGCCTT	949
Qy	1305	gctgagtcgacagaaatccctgagccctactggaatgtataatgtgttggcagct	1364
Db	950	GCAGAGCAGAAAATCAGAACTGGATTTTCAGTACTGGAACCTTCAATGTGTAGCTGTT	1009
Qy	1365	aaaattgctgttctgctccagttgacagatcccaagccctacgtataacatggaacttaatt	1424
Db	1010	AGAAATCGCTGTCTGGCATCCATTTGGTTACTCAGGCATTTATGATGTGGAATTCATT	1069
Qy	1425	actctcgttcagaggtgggtagaagattctaattcagccctcatgtatgaaaaag	1484
Db	1070	AATTTCAGCTCTCAGAGGTGGAGGACATCTCTGCTTTTCAGGCACCAGCTGTGGAAGAAG	1129
Qy	1485	aaacggctc-----gagatctcttaaaaaaagacagaaacggagtg---gga	1529
Db	1130	AAACCAACAGTAACTAAGGCAGATCTCTTAAAAAAGAACAGAAAATGTGTGTAATGGA	1189
Qy	1530	gtggaacttcaaatagatagactgtccgcaagaggaaagagaataatttcataat	1587
Db	1190	ACATTAACTTCAAAATGTAGCAGACTCTCCCGGGAATAAAAAAGAAAAATCTTCATAAT	1247
RESULT 11			
AY029764			
LOCUS			
DEFINITION Mus musculus TRAM1 mRNA, complete cds.			
ACCESSION AY029764			
VERSION AY029764.1 GI:13774094			
KEYWORDS house mouse.			
SOURCE			

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1. (bases 1 to 2720)
TITLE	Hartmann, E.
JOURNAL	Direct Submission
Allee 160, Luebeck 23538, Germany	Submitted (09-APR-2001) Biologie, Universitaet Luebeck, Ratzeburger
FEATURES	Location/Qualifiers
source	1. .2720
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="R2PD clone IMAGp998K219524q2"
CDS	36. .1160
	/note="Tram1"
	/codon_start=1
	/product="TRAM1"
	/protein_id="AAK38167.1"
	/db_xref="GI:13774095"
	/translation="MAIRKSNKPPLLSHEFLQNHADIVSLAMFLILGLMFVETA KGAIIFVALQYNTWPATEQATSESLYHGKIDLATVLFYMLVAIIHIIQEVVL KNIRRMHESKTKHSPNESGOLSAFLFACVWGFTILISENYISDPTILIRAYPHNL MTQNRFFYSIQLAYLWHPPEIDYFQNTKKEDIPLRVYIGLIFHIAIGAYLLNLNL GLVLLVHYFVELEFHSIRLFYSEKQKFSLWALFVLGRLTLILSLVTVGFL ARAENOKLDFSTGNFNLAVRIAVLASICITQAFMMKFINFOLRWRHESAFQAPPV KRPVATKGRSSRKGTENGVTGTVTSNGADSPNRKES"
BASE COUNT	722 a 569 c 530 g 899 t
ORIGIN	
Query Match	32.0%; Score 733; DB 10; Length 2720;
Best Local Similarity	74.0%; Pred. No. 2.3e-166;
Matches 1005; Conservative	0; Mismatches 330; Indels 24; Gaps 5;
Qy	465 ggaagcgccaccatgggggtccctgaagaagcagcaccacccccccgtctcagccag 524
Db	24 GCGGGCTCCACCTAGCGGATTCGCAAGAGCAAGCAACGCCGCTGCTGAGCCAC 83
Qy	525 gaattcatcctcagaatcagcgacatcgctctcgtgggagttcttctcctgctg 584
Db	84 GAATCTTGTCTGAGAAATCAGCGGACATCGTCTCTGCTGCGGATGCTTTCTCTGCTG 143
Qy	585 gggctgtgtctgaggaagcagcagcagcagcagcagcagcagcagcagcagcagc 644
Db	144 GGCTTATGTTTCAGGTAAACAGCAAAAGGAGCCATCATCTTTGTGCTCTTCAGTAAAT 203
Qy	645 gtgctgtccctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 704
Db	204 GTTACCGCGCTCAACAGAAAGCAAGCTACTGAATCAGCATCCCTATATCATATAGGT 263
Qy	705 gtcaaaagatttggccagcgtttctctctacatgctggtggcaatcattatcagccaca 764
Db	264 ATCAAGACTTGGCTACAGTTTGTCTACATGCTAGTGGCGATAAATTATTCATGCCATA 323
Qy	765 attcaggaatatgtgttgataaaattaaacaagaagaatgcagttccacaaagcgaacaa 824
Db	324 ATTCAAGAGTATGTTGGATAAAATTAACCGAGGATGCATTTCTCCAAAACGAAGCAC 383
Qy	825 acaagtttaacagcagctcgtgctgagtttagtgtgttctacttttttttttttttttttt 884
Db	384 AGCAAGTTTAAACGAGCTGTGCTAGCTCAGTGCATTTTACCTTTTGGCTGTGTGGGGC 443
Qy	885 acattcattttaaactctgtaaaactgcctgcagaccacaaactcttatatggaaggtcgt 944
Db	444 ACCTTCATCTCCTCCGAAAACATATATCTCAGACCAACTATCTCAGGAGCCCTAT 503
Qy	945 ccccatagcatgatgacatttcaaatgaagtttttttttttttttttttttttttttttttt 1004
Db	504 CCCCATAACTGTGATGACATTTTCAGACGAAGTTTTTCTTACATCTCCAGCTGGGTACTGG 563
Qy	1005 ttctatgcttttctgaactctacttccagaaacccaaacaaacaaacaaacaaacaaacaa 1064
Db	564 CTCCATGCATTCCTGAACTCTACTTCCAGAAAACCAAAAAGAAATATATCCCTCGTCAG 623

QY	1065	cttgctacattggtgttccaaactcttccaaacttactgagagcttatctctgttaacttgaat	1124
Db	624	CTCTCTACATGTTGCCCTTATCTCTTTTCATATCTCTGAGCTATCTCTCTGAACCTTGAAC	683
QY	1125	catttggaactctctcttgggtactcattatttggttgaattacttctcccaatgctgc	1184
Db	684	CACCTGGGCCCTTGCCCTTCCTGCTGCATGATATTTGCTTGAATTTCTTTTCCACATTTCT	743
QY	1185	ggcctgttttacttttagtgataaagtaccagaaaaggcatatctctgtggccattgtg	1244
Db	744	CGTCTATTTTATTTAGTGACGAGAAGTATCAGAAAAGGTTTTCCTCTGCGCAGTTCTT	803
QY	1245	tttatctgggtagaacttggtaactttaattgtttccctactactcaactgtttgggttccacctg	1304
Db	804	TTTCTTTTGGGAACATCTCTGACTCTAATCTTTTTCAGTTCTTACTGTGTTGTTTGGTCTC	863
QY	1305	gctgagtcgcagaatccgaatccctgagtcgcccttactgagaatgtaaatgtgtttgcagct	1364
Db	864	GCAAGAGGGAGAAATCAGAACCTGACTTTAGTTAGTTGTAACCTTCAACGCTGTGGCCGTT	923
QY	1365	aaaattgctgtctgtccagttgacgagtcacagctcgaagctcagtaacatgaaacttaatt	1424
Db	924	AGGATCGCTGTTCTCGCGTCCATTGGCATCACACAGCTTCATGATGTGGAAGTTCAATT	983
QY	1425	actctctggctcagaggtgggtagaagattcttaattatcaggccctcatgtatgaaaaag	1484
Db	984	AAC'TTCCAGCTTCCGAGGTGGAGGGAACATCTGCCTTCCAGGCCCACTGTGAAAAGG	1043
QY	1485	aaacggtc-----gagatcttctaaaaaaagacagaaaacgagagtg-----gga	1529
Db	1044	AAACCGGCGGTGACCAAGCGAGCTGTCCAGAAAAGGACAGAGACGGTGTGAACGGC	1103
QY	1530	gtggaacttcaaatagtagtagactgtccgcgaagaggaagaaatctctcataatct	1589
Db	1104	ACACTACATCAACGAGGAGACACTCCCTCGGACAGGAAGAGAGACTCTTCGTAATGG	1163
QY	1590	ttgcaagcgcaattgatttaattgtctgcaaggaaatctctcttttgggtttctttctgcac	1649
Db	1164	ATCGGAAGCTGACTGATATTGTTGCCCAAGAAATCTGCTTCTCTATCTTTTCAGCAC	1223
QY	1650	tagagattttctgtttttgaaa--atagttctgtctctctctgtttttgttattgaactgt	1708
Db	1224	TAGCAATTTTCTTTTCTAAAAATACAGTTTATCTCTTGTGATTTCTGCTGTTGCTGCT	1283
QY	1709	tctatgattttttaaagacatttgaggggagggagattattatgaatgggaaaaaaga	1768
Db	1284	TCC---TGCATGCTAAAGGCGATTGAGGGGAGGATAATCACTCTGAGTG-----AGAACA	1335
QY	1769	ttttggttgagactaaattactcatctcgtcataaataatgt	1807
Db	1336	TTGTAGCTTAGACTCAGCTACCTACCTGCCCTCAGAAATAGTTT	1374
RESULT	12		
LOCUS	BC012401		
DEFINITION	Mus musculus, clone MGC:11724 IMAGE:3967323, mRNA	2819 bp	linear ROD 20-AUG-2001
ACCESSION	BC012401		
VERSION	BC012401.1	GI:15214554	
KEYWORDS	MGC.		
SOURCE	house mouse		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 2819)		
JOURNAL	Direct Submission		
	Strausberg, R.		
	Submitted (15-AUG-2001), National Institutes of Health, Mammalian		
	Gene Collection (MGC), Cancer Genomics Office, National Cancer		
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,		
	USA		
REMARK	NIH-MGC project URL: http://mqc.nci.nih.gov		

Qy	825	aacaagttaa	cgagctggtcagtttagt	gtgttctactttttttctgttatttgggc	884
Db	481	ACGAAGTT	TAAGAGCTGGTCAGCTAGTCATCTACCTTTTGCCTGTGTGGGCG	540	
Qy	885	acattcattt	aaatctctgaaacgcctgtcagacccaactcttatggaaggtcgt	944	
Db	541	ACGTTCA	TCTCATCTCCGAAACATATATCTCAGACCCCAACTATCTGTGGAGACCTAT	600	
Qy	945	ccccatagc	atgataacatttcaaatgaattttctacatatccccagttggcttactcg	1004	
Db	601	CCCCATAC	CTGATGACATTTTCAGACGAAGTTTTCATCTCCACCTGGCGTACTGG	660	
Qy	1005	tttcatgctt	cttctgaaactctacttccagaaaaacaaaacagacatccctcgtcaa	1064	
Db	661	CTCCATG	ATGATCCCTGACTCTACTTCCAGAAACCAAAAGAGATATCCCTCGTCAG	720	
Qy	1065	cttgtctac	atggtcttcaactcttccacattactcagagcttatctcttctgactgaat	1124	
Db	721	CTCGTCA	TACATTTGGCCTTTATCTTTTCATATTTGCTGGAGCCTATCTTCTGAAC	780	
Qy	1125	catctggaa	ctctcttcttggtaactcattattttgttgaaattactttccacatgtgc	1184	
Db	781	CACCTGG	CCCTGCTGCTGCTGCTACTATTTCTGTGAATTTCTTTTCCACATTTCT	840	
Qy	1185	ggcctgttt	tacttttagtgataaagaccagaaagccatctctcgtggccattgtg	1244	
Db	841	CGTCTAT	TTTATTTAGTGACGAGAAGTATCAGAAAGGGTTTCTCTCTGGCAGTTCTT	900	
Qy	1245	tttatctgg	tagactgtgactttaattgtttccgtactcactgttgggtttcaactg	1304	
Db	901	TTTGTGT	TGGGAAGACTTCTGACTCTAATCTTTTCTCAGTCTTACTGTGTGTTGTC	960	
Qy	1305	gctggatcg	cagaatcctgactgccttactgaaatgtaattgtttggcagct	1364	
Db	961	GCAAGCG	GGAGAAATCAGAAGCTGGACTTTAGTACTGGAAATCTCAACGTGTGGCGTT	1020	
Qy	1365	aaaattgt	ctgtctgtcgtccagttgacgcatccaaagccctacatgaactgaaacttaatt	1424	
Db	1021	AGGATCG	CTGTCTCGCGTCCATTTGCATCACACAAGCCTTCATGATGTGGAAGTTCA	1080	
Qy	1425	actctcgt	ctcagagtggtgtagaagattctaatatctcaggtcctcattgataaaaaag	1484	
Db	1081	AATTTCC	AGCTCGGAGTGGAGGGAACATTTCTGCTTTCCAGGCCCCACCTGTGAAA	1140	
Qy	1485	aaacggctc	-----gagatcttctaaaaaagacagaaaaacggagtg---gga	1529	
Db	1141	AAACCGG	CCGTGACCAAGGAGGCTCTTCCAGAAAAGAACAGAAAGGTTGTAACGGC	1200	
Qy	1530	gtggaaact	caaatagatgactgtcgcgaagagagaaatcttccataatct	1589	
Db	1201	ACAGTCA	CATCAACAGGAGCAGACTCCCTCGGAACAGGAAAGAGTCTTCGTAATGG	1260	
Qy	1590	ttgcaagc	gattgaattatctgcaagaatctgctcttggaggtttcttcttctgcac	1649	
Db	1261	ATCGGA	AGCTGACTGATTTATTTGTCGCCAAAGAAATCTGCTTTCTATCTATCTC	1320	
Qy	1650	tagagattt	ctgttttgaaa-atagttcgtcgtctcgtgtttttgttattgaactgt	1708	
Db	1321	TAGCAAT	TTTTTCTTAAATAATACAGTTATGCTCTTTGATTTCTGCTGTGTGCTGT	1380	
Qy	1709	ttcatgtat	tttttaagacattttgaggggaggagattattatgaattgggaaaaaaga	1768	
Db	1381	TCC---	TGCATCGTAAAGGCGATTGAGGGAGGATTAATCACTCTGAGTG-----	1432	
Qy	1769	ttttggtg	agactaaataactcactcgtcgaataatgt	1807	
Db	1433	TTGTAGC	TTAGACTACGCTACCTGCTGCTTTCAGAAATAGTTT	1471	

RESULT 13

AC027246/c

LOCUS

HTG 28-MAR-2000

73339 bp DNA linear

AC027246

DEFINITION	Homo sapiens chromosome 4 clone RP11-724E21 map 4, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC027246
VERSION	AC027246.1 GI:7331616
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Homo sapiens chromosome 4, clone RP11-724E21
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 73339)
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavsky,I., Boukhgalter,B., Brown,A., Burkett,G., Campiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A., Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (28-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
	Center: Whitehead Institute/ MIT Center for Genome Research
	Center code: WIBR
	Web site: http://www-seq.wi.mit.edu
	Contact: sequence.submissions@genome.wi.mit.edu
	Project Information
	Center project name: L8852
	Center clone name: 724_E_21

	* NOTE: This record contains 82 individual
	* sequencing reads that have not been assembled into
	* contigs. Runs of N are used to separate the reads
	* and the order in which they appear is completely
	* arbitrary. Low-pass sequence sampling is useful for
	* identifying clones that may be gene-rich and allows
	* overlap relationships among clones to be deduced.
	* However, it should not be assumed that this clone
	* will be sequenced to completion. In the event that
	* the record is updated, the accession number will
	* be preserved.

	* 1 775: contig of 775 bp in length
	* 776 875: gap of 100 bp
	* 876 1673: contig of 798 bp in length
	* 1674 1773: gap of 100 bp
	* 1774 2552: contig of 779 bp in length
	* 2553 2652: gap of 100 bp
	* 2653 3464: contig of 812 bp in length
	* 3465 3564: gap of 100 bp
	* 3565 4277: contig of 713 bp in length
	* 4278 4377: gap of 100 bp
	* 4378 5176: contig of 799 bp in length
	* 5177 5276: gap of 100 bp

```

* 5277 6071: contig of 795 bp in length
* 6072 6171: gap of 100 bp
* 6172 6952: contig of 781 bp in length
* 6953 7052: gap of 100 bp
* 7053 7844: contig of 792 bp in length
* 7845 7944: gap of 100 bp
* 7945 8722: contig of 778 bp in length
* 8723 8822: gap of 100 bp
* 8823 9633: contig of 811 bp in length
* 9634 9733: gap of 100 bp
* 9734 10513: contig of 780 bp in length
* 10514 10613: gap of 100 bp
* 10614 11422: contig of 809 bp in length
* 11423 11522: gap of 100 bp
* 11523 12315: contig of 793 bp in length
* 12316 12415: gap of 100 bp
* 12416 13201: contig of 786 bp in length
* 13202 13301: gap of 100 bp
* 13302 14093: contig of 792 bp in length
* 14094 14193: gap of 100 bp
* 14194 14978: contig of 785 bp in length
* 14979 15078: gap of 100 bp
* 15079 15877: contig of 799 bp in length
* 15878 15977: gap of 100 bp
* 15978 16781: contig of 804 bp in length
* 16782 16881: gap of 100 bp
* 16882 17669: contig of 788 bp in length
* 17670 17769: gap of 100 bp
* 17770 18566: contig of 797 bp in length
* 18567 18666: gap of 100 bp
* 18667 19473: contig of 807 bp in length
* 19474 19573: gap of 100 bp
* 19574 20373: contig of 800 bp in length
* 20374 20473: gap of 100 bp
* 20474 21273: contig of 800 bp in length
* 21274 21373: gap of 100 bp
* 21374 22175: contig of 802 bp in length
* 22176 22275: gap of 100 bp
* 22276 23074: contig of 799 bp in length
* 23075 23174: gap of 100 bp
* 23175 23965: contig of 791 bp in length
* 23966 24065: gap of 100 bp
* 24066 24866: contig of 801 bp in length
* 24867 24966: gap of 100 bp
* 24967 25772: contig of 806 bp in length
* 25773 25872: gap of 100 bp
* 25873 26685: contig of 813 bp in length
* 26686 26785: gap of 100 bp
* 26786 27577: contig of 792 bp in length
* 27578 27677: gap of 100 bp
* 27678 28470: contig of 793 bp in length
* 28471 28570: gap of 100 bp
* 28571 29373: contig of 803 bp in length
* 29374 29473: gap of 100 bp
* 29474 30223: contig of 750 bp in length
* 30224 30323: gap of 100 bp
* 30324 31114: contig of 791 bp in length
* 31115 31214: gap of 100 bp
* 31215 32038: contig of 824 bp in length
* 32039 32138: gap of 100 bp
* 32139 32951: contig of 813 bp in length
* 32952 33051: gap of 100 bp
* 33052 33845: contig of 794 bp in length
* 33846 33945: gap of 100 bp
* 33946 34753: contig of 808 bp in length
* 34754 34853: gap of 100 bp
* 34854 35662: contig of 809 bp in length
* 35663 35762: gap of 100 bp
* 35763 36571: contig of 809 bp in length
* 36572 36671: gap of 100 bp
* 36672 37457: contig of 786 bp in length
* 37458 37557: gap of 100 bp
* 37558 38344: contig of 787 bp in length

```

```

* 38345 38444: gap of 100 bp
* 38445 39246: contig of 802 bp in length
* 39247 39346: gap of 100 bp
* 39347 40118: contig of 772 bp in length
* 40119 40218: gap of 100 bp
* 40219 41012: contig of 794 bp in length
* 41013 41112: gap of 100 bp
* 41113 41915: contig of 803 bp in length
* 41916 42015: gap of 100 bp
* 42016 42808: contig of 793 bp in length
* 42809 42908: gap of 100 bp
* 42909 43708: contig of 800 bp in length
* 43709 43808: gap of 100 bp
* 43809 44604: contig of 796 bp in length
* 44605 44704: gap of 100 bp
* 44705 45507: contig of 803 bp in length
* 45508 45607: gap of 100 bp
* 45608 46403: contig of 796 bp in length
* 46404 46503: gap of 100 bp
* 46504 47319: contig of 816 bp in length
* 47320 47419: gap of 100 bp
* 47420 48211: contig of 792 bp in length
* 48212 48311: gap of 100 bp
* 48312 49111: contig of 800 bp in length
* 49112 49211: gap of 100 bp
* 49212 49925: contig of 714 bp in length
* 49926 50025: gap of 100 bp
* 50026 50816: contig of 791 bp in length
* 50817 50916: gap of 100 bp
* 50917 51719: contig of 803 bp in length
* 51720 51819: gap of 100 bp
* 51820 52609: contig of 790 bp in length
* 52610 52709: gap of 100 bp
* 52710 53505: contig of 796 bp in length
* 53506 53605: gap of 100 bp
* 53606 54387: contig of 782 bp in length
* 54388 54487: gap of 100 bp
* 54488 55295: contig of 808 bp in length
* 55296 55395: gap of 100 bp
* 55396 56184: contig of 789 bp in length
* 56185 56284: gap of 100 bp
* 56285 57082: contig of 798 bp in length
* 57083 57182: gap of 100 bp
* 57183 57992: contig of 810 bp in length
* 57993 58092: gap of 100 bp
* 58093 58890: contig of 798 bp in length
* 58891 58990: gap of 100 bp
* 58991 59775: contig of 785 bp in length
* 59776 59875: gap of 100 bp
* 59876 60669: contig of 794 bp in length
* 60670 60769: gap of 100 bp
* 60770 61575: contig of 806 bp in length
* 61576 61675: gap of 100 bp

```

```

Query Match      28.3%; Score 648.2; DB 2; Length 73339;
Best Local Similarity 99.5%; Pred. No. 1.8e-145;
Matches 650; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1636 gttctttctgcactagagattttctgtttttgaaaaatagtcgtcgtcgtgtttt 1695
DB 71476 GGTTCCTCTGCACTAGAGATTTTCTGTTTTTGAATAAGTTCGTGCTCTCGGTTTT 71417
QY 1696 gttattgaactgttcattgtatttttaagacatttggggagagattattatgaa 1755
DB 71416 GTATTGAACTGTTTCATGTATTTTTTAAGACATTTTGAAGGAGGAGGATTTATGAA 71357
QY 1756 tgggaaaaaagattttggttgagactaaattactcatcgtcaaaaataatgtcaaatag 1815
DB 71356 TGGGAAAAAAGATTTTGCTTGAGACTAAATTTACTCATCTGCAAAATAATGTCAAAATAG 71297
QY 1816 ttttggggatcacccactatattttgttggatttttaacctttcaacatttctctaata 1875
DB 71296 TTTTGGGGATCACCACTATATTTTGTGTTTTTAACTTTTAACTTTTCAACATTTTCTCTAATGA 71237

```

```

QY 1876 ttgcagataaactgcacaattttgcatatcaatgatgaactggtttactccacacagt 1935
Db 71236 TTTGCAGAGATAAAGTGCACAAATTTGCATATCAATGATGACTGGTCTACTTCCACACAGT 71177
QY 1936 gtttcataataactaacagatggtctctctcctagcaagattatggtttaatgcttgcgttt 1995
Db 71176 GTTTCATAATACTAACAGATGGTCTCTCTCTAGCAAGATTATGTTTAAATGCTGCTTT 71117
QY 1996 ggggttaataaaagtcagaaaaagtggaagtcacaaatcagatctctctgtaattgtagaa 2055
Db 71116 GGGGTAAATAAAGTAGCAAAAAGGTGGGAAGTCAAAATCAGTATCTGTAAATTTGTAGAA 71057
QY 2056 tttatttttaagaacttaacaactcagaaaaagattgctagactcaccacaaataataaattg 2115
Db 71056 TTTATTTTAAAGAACTTACAACCTCAGAAAAGATTGCTAGACTCACCAGAAATTAATAATG 70997
QY 2116 tctttattttacaggtagtgattatttagtgcttcattccaccatttaaaaaaacacagtac 2175
Db 70996 TTCTTTATTTTACAGGTAGTGATATTAGTGTCTCATCCCAATTTAAAAAACACAGTAC 70937
QY 2176 taatgggtaacacatatggaggttgcgcacatatattgctgcacataattgcatcaaaatacattaa 2235
Db 70936 TAATGGGTAAACATATGGAGGTTTGTGCGCATATATATGCAATCAAAATATCATTAAT 70877
QY 2236 aatataaaaataaaaacttcctctgctccattccaccattgtaaatgggaattc 2288
Db 70876 AATATAAAAATAATAAACTTTTCTGTCTCATCCCTTGTAAATGGGAATTC 70824

RESULT 14
BD005227
LOCUS BD005227 1092 bp DNA linear PAT 31-JAN-2002
DEFINITION Method of screening of protein.
ACCESSION BD005227
VERSION BD005227.1 GI:18633188
KEYWORDS JP 03075332-T/2.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1092)
Todo,N., Okuyama,H., Imamura,M., Ishikawa,H. and Nemoto,K.
Method of screening of protein
Patent: JP 03075332-T 2 16-FEB-2001;
SUMITOMO PHARMACEUTICALS CO LTD,NAOKI TODO,HAJIME OKUYAMA, OTOAKI
IMAMURA, HIRONORI ISHIKAWA, KIYOMITSU NEMOTO
OS Rattus sp. (rat)
PN JP 03075332-T/2
PD 16-FEB-2001
PF 17-AUG-2000 JP 2000005488
PI 20-AUG-1999 JP 99P 234764
PI NAOKI TODO,HAJIME OKUYAMA,MOTOAKI IMAMURA,HIRONORI ISHIKAWA,
PI KIYOMITSU NEMOTO
PC C1261/02,G01N33/50,C07K14/47,A61K38/17,C12N5/10,C12P21/02// PC
(C12P21/02,C12R1:91)
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..1092
/organism="Rattus sp."
/db_xref="taxon:10118"
BASE COUNT 260 a 268 c 288 g 276 t
ORIGIN

Query Match 28.3%; Score 647.8; DB 6; Length 1092;
Best Local Similarity 76.1%; Pred. No. 7.5e-146;
Matches 844; Conservative 0; Mismatches 247; Indels 18; Gaps 3;

QY 477 atggggctcgttaagaagagaccacaaagacccccccggttctcagcagggaattcctg 536

```

```

Db 1 ATGGGGCTCGCAAGAAGAACGCGCAGAAACCCCGGTCGTCAGCCACGAAATTCATGGTG 60
QY 537 cagaatacatcggaacatgctctcctcgtcgtgggagatcttctcctcgtggggtctgtgttc 596
Db 61 CAGAACACGCGGATATGGTCTCTCGTGGGCATGTTCTGCTGGGACTTATGTTTC 120
QY 597 gggggaacagcagaagcaaccatcgtgttctcactcttcacacagatgttctgctccct 656
Db 121 GAGGGCACGCGCGAGATGCGATCGTGTTCCTCACCTCGACATGGAGTGGATGTTGTC-- 178
QY 657 gcagcagaagaacacagccagcgggtccaaagtccctctattattatgggtgcaaaagtattg 716
Db 179 -CAGCGGAAGGGCTACCCCTCGGGGTCCAGAGACCTTTACCATATTGGGGTCAAGATCTG 237
QY 717 gccacggtttctctacatgctggtggcaatcattattcattcgcacacaaatcagaatat 776
Db 238 GCCACAGTGTCTTCTACATGCTGGTGGCCATCATCTTCAGCCACCATTCCAGGAGTAC 297
QY 777 gttgtggataaaattacaagaagaatgcagttccacaaagcgaaacacaaagtttaac 836
Db 298 GTGCTAGATAAGCTCAGCGGGAGACTGCAGCTCACCAAGGCAACAAACAATTTGAAT 357
QY 837 ggtctggtcaggtttagtgatttctacttttttttttttttttttttttttttttttttt 896
Db 358 GAGGCGGGCAGCTGAGTGTCTTACATAGTGTCTGTGTATCTGGGTATGATCATCTCTG 417
QY 897 atctctgaaaactgcctgcagaccccaactcttatatggaaggtcgtcccccacagatg 956
Db 418 GCCTCTGAGAACTGCCTCTCAGACCCCACTCTATTGTGGAAGTCTCAGCCCCCAACAATG 477
QY 957 atgacatttcaaatgaagttttctacatatccacagttggcttactggtttcactgctttt 1016
Db 478 ATGACATTTCAGATGAATTTTCTACATCTCAGAGTGGCTTACTGTTGTTTCATAGTTTC 537
QY 1017 cctgaactctacttccagaaaaacacaaacacacacacacacacacacacacacacacac 1076
Db 538 CCGGAGCTCTACTTCCAGAAAGTCAGGAACAGATATCCCGGGTCAACTCATCATATT 597
QY 1077 ggtctcacctctccacattactgagcttatctctgtacttgtaactgaatcattggagct 1136
Db 598 GGCTCCACCTCTTCCACATTGGAGGGCCCTATCTCTGTACTTGAACACCTGGGCGTG 657
QY 1137 cttctttgttactgcatattttgttgtaattacttccacacatgctggcgcctgttttac 1196
Db 658 CTGCTTCTGATGTGCACTATGCTGTGAGCTCTCTCCAGCTCTCTCCAGCTGTGTCAGCTT 717
QY 1197 tttagtgatgaaaagtaccagaaggcatatctctgtgggcccattgtgtttacttctgggt 1256
Db 718 TTTGGGATGAGCGGTACCAGAAAGGTTGCTTTGTGGCTATCGTGTATATATCCGGG 777
QY 1257 agacttgtaactttaattgttccgtactcactggttgggtttcacctggctggatgcag 1316
Db 778 AGACTCGTACACTGATGTGTCTCAGTGGTGTACAGTGGGCTTCACTTGGCCGG---GACA 834
QY 1317 aatcggaactcgtgatcccttactggaagtgaatgtgttgagcagctaaattgctgtt 1376
Db 835 AATCGGAATGGAATGCTCTCTCTGTTATGTCATGTGTGGACCTAAAAATCGCTGTT 894
QY 1377 ctgtctccagttgcacgatccaagccttaacatcgtaactgaacttaattactctctgctt 1436
Db 895 CTGTCTCGAGTTGCAGTATCCAGGTGTACATAACATGAGCTTGACGACCCCTGCGCTT 954
QY 1437 cagagtggttagaagattcttaatttcaggcctcatgtatgaaaaaagaaacggtcgaga 1496
Db 955 CAGAGATGTTTAGAAGATCGGAATCTTCTATGTTCTGTGGAGGAA-----GAGA 1002
QY 1497 tcttcaaaaaagacagaaacgagtgagggtggagtggaacttcaaatagatgactgt 1556
Db 1003 CGGTCCAGTCCAGAAAAGGCCACAGAAAATGGAGTGGAGAAATCCAAATAGATAGATTCT 1062
QY 1557 ccgccaagaggaagaaagaaatcttcata 1585

```

Db 1063 CCACCAAGAAGAAAGAGAAAGCTCCTTA 1091

RESULT 15

AY029534 1338 bp mRNA linear VRT 02-MAY-2001
LOCUS
DEFINITION
Xenopus laevis Traml mRNA, complete cds.
ACCESSION
AY029534
VERSION
AY029534.1 GI:13936286
KEYWORDS
SOURCE
African clawed frog.
ORGANISM
Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 1338)

REFERENCE

Hartmann, E.
Direct Submission
Submitted (10-APR-2001) Biologie, University Lubeck, Ratzeburger
Allee 160, Lubeck 23538, Germany

FEATURES

Location/Qualifiers
1..1338
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="RzPD A01802Q2"
16..1137
/codon_start=1
/product="Traml"
/protein_id="AAK40302.1"
/db_xref="GI:13936287"

CDS

translation="MGIRKKNKSKTPPVLSHEFIIONHADIVISCLAMVFLGLMFBEITA
KVAVMFVLQNVTPVEGVLCSESLYHGIKDMATVFFYMLVAIILHAVIOEYVLD
KINRMHFSKTKHSENGSOLFSLFSCIWGASIIIVSENVFSDPISLWKGYPHTYF
PNORKEFFISQLAYNFHAPPELFOFKKEDIPROLVIGLYLFIHIGAYVLNLRGL
LYLLVHLIFVFLFMSRDFYFSNERYQGFVMAVFLVGLRLFLIILSVLTVGFLA
RAENQELDLNNGNFIILAIRITVLASICITAFNMKFINQLRWRHREHSSPPQSSOR
KKATSAKGAASKKXENGVNGVTVTNSGADSPRSRKEKS"
BASE COUNT 362 a 321 c 267 g 388 t
ORIGIN

Query Match 22.2%; Score 508.8; DB 5; Length 1338;
Best Local Similarity 67.7%; Pred. No. 3.3e-112;
Matches 765; Conservative 0; Mismatches 347; Indels 18; Gaps 3;

QY 472 ccacatggggccgtcgaagaagaccacgaaccccccccggtttctcagccaggaaattca 531
Db 11 CCACCATGGGCATCCGTAAAGAAAGACGACGACCCCGCGGTGCTCAGCCACGAGTTCA 70
QY 532 tctgcagaatcatcgccgacatcgtctcgtggtggggtgtcttctcgtggtgggcttg 591
Db 71 TTATCCAGAACCATCGGACATCGTCTCTCGCTGGCCATGGTCTTCTGCTCGGCCTTA 130
QY 592 tgttcgagggaaacagcagcagcatcctcgtgtttctcactcttcagcagcagtggtgctg 651
Db 131 TGTTCAGATTACAGCAAAAGGTGGTGTCTATGTTTGTACACTCCAGTACAAATGTCAACCA 190
QY 652 tccctgcacagagaacacgaacggctcaagtcctctctattattatggtgtcaaaag 711
Db 191 TTCCAGTAGAAGGGGTGCTTG---GAGAACCCACATTCGCTGTACCATTTATGGCATCAAAG 247
QY 712 atttgccacgggtttctctctacatcgtgttggaatcattatcattgcacaaattcagg 771
Db 248 ATATGCCCACTGATTTTTTACATGCTGTGGCATATATACTACATGCAAGTTATCCAAG 307
QY 772 aatatgtgttgataaaaaataaacaagagaatgcagttccacaaagcgaaacaaacaagt 831
Db 308 AATACGTATTAGATAAAATCAACCGCGGTATGCATTTTCCAAAGACAAAAACACAGCAAGT 367
QY 832 ttaacagagctcgtcagtttagtgttctcactttttttcttctgtatttggggcacattca 891
Db 368 TTAATGAGTCTGGACAGCTAAAGTGCATTTTACCTGTCTCTTGTGATTTGGGGAGCAAGCA 427

Search completed: September 6, 2002, 14:51:47
Job time: 36711 sec

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 14:57:08 ; Search time 495.6 seconds
(without alignments)
7926.352 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcgcgcgtgtgcga.....ccactgttaaatggaattc 2288

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

```

1: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2288	100.0	2288	21 AAA38013	Human WAR-1 nucleo
2	1808.8	79.1	1835	23 AAS76421	DNA encoding novel
3	1734.4	75.8	1736	22 AAK94181	Human full-length
4	1348.2	58.9	1517	24 AAS18576	cDNA encoding huma
5	1288	56.3	1288	22 AAF74782	Human WAR-1 encodi
6	1110	48.5	1110	22 AAF74780	Human WAR-1 encodi
7	937.6	41.0	2833	21 AAF77810	Human cancer assoc
8	923.4	40.4	2756	24 AAS62269	cDNA sequence #56
9	802.2	35.1	2311	21 AAA38012	Rat WAR-1 nucleoti

10	761.6	33.3	802	22	AAK91674	Human cDNA 5'-end
11	761.6	33.3	802	22	AAK93837	Human cDNA clone r
12	679.2	29.7	1120	20	AAZ33531	Human prostate can
13	647.8	28.3	1092	22	AAZ74781	Rat WAR-1 encoding
14	504.4	22.0	545	22	AAK92492	Human cDNA 3'-end
15	486.8	21.3	642	22	AAZ74389	Human breast cance
16	486.8	21.3	642	22	AAZ17959	Human breast cance
17	298.8	13.1	1380	22	AAZ01678	Human reproductive
18	273	11.9	472	21	AAZ00195	Human secreted pro
19	161	7.0	1526	23	ABL18595	Drosophila melanog
20	161	7.0	1549	23	ABL18061	Drosophila melanog
21	161	7.0	2153	23	ABL21096	Drosophila melanog
22	161	7.0	3346	23	ABL21098	Drosophila melanog
23	161	7.0	3618	23	ABL18060	Drosophila melanog
24	161	7.0	4146	23	ABL18594	Drosophila melanog
25	161	7.0	4775	23	ABL21091	Drosophila melanog
26	161	7.0	5055	23	ABL21094	Drosophila melanog
27	161	7.0	11287	23	ABL21090	Drosophila melanog
28	119	5.2	447	22	AAZ28950	Colon tumour relat
29	93.8	4.1	271	22	AAZ23061	Human breast cance
30	92.2	4.0	205	16	AAZ19252	Human gene signatu
31	91.4	4.0	272	22	AAZ14199	Human breast cance
32	88.4	3.9	494	24	ABZ99445	Mouse ischaemic co
33	85.6	3.8	349	21	AAZ06201	Human secreted pro
34	65.6	2.9	602	22	AAZ60814	Human cancer agent
35	58.4	2.6	3641	23	AAZ88713	DNA encoding novel
36	58.4	2.6	6475	23	AAZ91168	DNA encoding novel
37	58.4	2.6	10771	23	AAZ85906	DNA encoding novel
38	54.8	2.4	6478	22	AAZ45416	Chemically pretrea
39	54.8	2.4	6626	22	AAZ46809	Tumour suppressor
40	52.6	2.3	7498	24	ABL32257	Human immune syste
41	52.4	2.3	16766	24	ABL34157	Human immune syste
42	52.2	2.3	6106	22	AAZ46430	Tumour suppressor
43	52.2	2.3	6106	22	ABL33473	Human immune syste
44	52	2.3	15416	24	ABL34231	Human immune syste
45	52	2.3	15416	24	AAZ61453	Human gene regulat

ALIGNMENTS

```

RESULT 1
AAA38013
ID AAA38013 standard; DNA; 2288 BP.
XX
AC AAA38013;
XX
DT 22-AUG-2000 (first entry)
XX
Human WAR-1 nucleotide sequence.
XX
DE Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
XX
KW diagnosis; cancer; sarcoma; human; ds.
XX
OS Homo sapiens.
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMO ) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX
PI Okuyama H;
XX
XX WPI; 2000-317980/27.
XX
XX P-PSDB; AAY98147.
XX
XX Endoplasmic reticulum protein WAR-1 which inhibits cancer cell

```


Db 1861 acatttcttaattgattgcagagataactgcacaatttttgcatatcaatgatctggtt 1920
QY 1921 cttactccaccagtggtttcataataactaacaagatggtctctcctcagcaagattatgtg 1980
Db 1921 cttactccaccagtggtttcataataactaacaagatggtctctcctcagcaagattatgtg 1980
QY 1981 ttaatgctgtcttgggtgtaaaataaaagtacgaaaaagggtggaagtcaaaatcagttat 2040
Db 1981 ttaatgctgtcttgggtgtaaaataaaagtacgaaaaagggtggaagtcaaaatcagttat 2040
QY 2041 cgtgaattgtagaatttttttaagaactacacactcagaaaaagattgctagactca 2100
Db 2041 cgtgaattgtagaatttttttaagaactacacactcagaaaaagattgctagactca 2100
QY 2101 ccaaaaataaaatgctttttttttacaggtagtgattattagtgcttcaccccatttt 2160
Db 2101 ccaaaaataaaatgctttttttttttacaggtagtgattattagtgcttcaccccatttt 2160
QY 2161 aaaaaaacagctactaattggttaacacatatggaggtttgtgcccataatatattgcac 2220
Db 2161 aaaaaaacagctactaattggttaacacatatggaggtttgtgcccataatatattgcac 2220
QY 2221 aaaaatcatttaataataaaaaattataaaatcattcctctgcccattccacttgtaaat 2280
Db 2221 aaaaatcatttaataataaaaaattataaaatcattcctctgcccattccacttgtaaat 2280
QY 2281 gggaattc 2288
Db 2281 gggaattc 2288

RESULT 2
AAS76421
ID AAS76421 standard; cDNA; 1835 BP.
XX AAS76421;
AC AAS76421;
XX
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #12225.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX P-PSDB; ABG12234.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
XX
XX Claim 1; SEQ ID No 12225; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, CC
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome mapping, and in recombinant production of (II). The CC

CC polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques CC
CC to restore normal activity of (II) or to treat disease states involving CC
CC (II). (II) is useful for generating antibodies against it, detecting or CC
CC quantitating a polypeptide in tissue, as molecular weight markers and as CC
CC a food supplement. (II) and its binding partners are useful in medical CC
CC imaging of sites expressing (II). (I) and (II) are useful for treating CC
CC disorders involving aberrant protein expression or biological activity. CC
CC The polypeptide and polynucleotide sequences have applications in CC
CC diagnostics, forensics, gene mapping, identification of mutations CC
CC responsible for genetic disorders or other traits to assess biodiversity CC
CC and to produce other types of data and products dependent on DNA and CC
CC amino acid sequences. AAS64197-AAS94564 represent novel human CC
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed CC
CC specification, but was obtained in electronic format directly from WIPO CC
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1835 BP; 533 A; 343 C; 356 G; 603 T; 0 other;

Query Match 79.1%; Score 1808.8; DB 23; Length 1835;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1810; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 477 atggggctccgtaagaagagcaccagaaccccccggttctcagccagggaattcactctg 536

Db 1 atggggctccgtaagaagagcaccagaaccccccggttctcagccagggaattcactctg 60

QY 537 cagaatcattcggaacatgctcctcgtggtgggatgttctctctgctgggctgtgttc 596

Db 61 cagaatcattcggaacatgctcctcgtggtgggatgttctctctgctgggctgtgttc 120

QY 597 gaggaacagcagaacatccatcgtgtttctcactctcagcacagtggtgtgcct 656

Db 121 gaggaacagcagaacatccatcgtgtttctcactctcagcacagtggtgtgcct 180

QY 657 gcagcagagagaacacccacgggctcaagtcctctattattatggtgtcaagattg 716

Db 181 gcagcagagagaacacccacgggctcaagtcctctattattatggtgtcaagattg 240

QY 717 gccacggtttcttctacatgctggtggcaatcattattcattcccaattcagggaat 776

Db 241 gccacggtttcttctacatgctggtggcaatcattattcattcccaattcagggaat 300

QY 777 ggttggtataaaattacaagaagaatgcagttccacaaagcgaacaaaacaaagttaac 836

Db 301 ggttggtataaaattacaagaagaatgcagttccacaaagcgaacaaaacaaagttaac 360

QY 837 gactctggtcagtttagtggtctctctttttctgtatttggggcacattcatttta 896

Db 361 gactctggtcagtttagtggtctctctttttctgtatttggggcacattcatttta 420

QY 897 atcttgaaaactgcctgcagacccaactcttatggaagcctgcctcccatagcatg 956

Db 421 atcttgaaaactgcctgcagacccaactcttatggaagcctgcctcccatagcatg 480

QY 957 atgacatttcaaatgaagtttttctacatccacagttggcttactggtttcattgttt 1016

Db 481 atgacatttcaaatgaagtttttctacatccacagttggcttactggtttcattgttt 540

QY 1017 cctgaacttacttccagaaaaacaaaacagacatccctcgtaactgtctacatt 1076

Db 541 cctgaacttacttccagaaaaacaaaacagacatccctcgtaactgtctacatt 600

QY 1077 ggtcttcaactcttccacattactgagcttatcttctgtacttgatcatttgggactt 1136

Db 601 ggtcttcaactcttccacattactgagcttatcttctgtacttgatcatttgggactt 660

QY 1137 cttcttttggtactgactgattttttgttgtaattacttccacatgtcgccgtgttttac 1196

Db 661 cttcttttggtactgactgattttttgttgtaattacttccacatgtcgccgtgttttac 720


```
QY 474 acatggggctcgttaagaagcagacccaagaacccccccggtttctcagccaggaattcctc 533
Db 181 accatggggctcgttaagaagcagacccaagaacccccccggtttctcagccaggaattcctc 240
QY 534 ctgcagaatcatgcggacatcctcctcgtcgtgggatgttcttctcgtcgtgggtctgtg 593
Db 241 ctgcagaaatcatgcggacatcctcctcgtcgtgggatgttcttctcgtcgtgggtctgtg 300
QY 594 ttccagggaaacagcagaatccatccatcgtgtttctcaactcttcagcacagtgctgtctc 653
Db 301 ttccagggaaacagcagaatccatccatcgtgtttctcaactcttcagcacagtgctgtctc 360
QY 654 cctgcagcagagaaacagccacgggctcaagtgccctctattattatgtgttcaagaat 713
Db 361 cctgcagcagagaaacagccacgggctcaagtgccctctattattatgtgttcaagaat 420
QY 714 ttggccacaggtttctctacatcgtgttgcaatcattattcattcagcccaattcaggaa 773
Db 421 ttggccacaggtttctctacatcgtgttgcaatcattattcattcagcccaattcaggaa 480
QY 774 tatgtgttgataaaattaaacagagaaatgcagttccacaaagcgaaacaaacaaagttt 833
Db 481 tatgtgttgataaaattaaacagagaaatgcagttccacaaagcgaaacaaacaaagttt 540
QY 834 aacgagctcgttcagtttagtgttctacttttttttttttttttttttttttttttttttt 893
Db 541 aacgagctcgttcagtttagtgttctacttttttttttttttttttttttttttttttttt 600
QY 894 ttaactctgaaactgcctgtcagaccacaaactcttatatggaaggctcgtccccaatgc 953
Db 601 ttaactctgaaactgcctgtcagaccacaaactcttatatggaaggctcgtccccaatgc 660
QY 954 atgatgacatttcaaatgaattttctctacatccagttggtctactggtttctatcgt 1013
Db 661 atgatgacatttcaaatgaattttctctacgtatccagttggtctactggtttctatcgt 720
QY 1014 ttctcgtgaactctctccagaaaaccccaaaacaaagacatccctcgtcaactgtctac 1073
Db 721 ttctcgtgaactctctccagaaaaccccaaaacaaagacatccctcgtcaactgtctac 780
QY 1074 attggtcttcaactctccacattactgaggttattctcttcttacttgaaatcatttgga 1133
Db 781 attggtcttcaactctccacattactgaggttattctcttcttacttgaaatcatttgga 840
QY 1134 ctctctcttttgactgactgattattttgtgaattacttccacatgtcgccgtcttt 1193
Db 841 ctctctcttttgactgactgattattttgtgaattacttccacatgtcgccgtcttt 900
QY 1194 tacttttagtgataaagttaccagaaaagccatctctgtggccattgtgtttatcttg 1253
Db 901 tacttttagtgataaagttaccagaaaagccatctctgtggccattgtgtttatcttg 960
QY 1254 ggtagaacttgtaacttaattgtttccgtactcaactgtgttggtttcaacctggtggtatcg 1313
Db 961 ggtagaacttgtaacttaattgtttccgtactcaactgtgttggtttcaacctggtggtatcg 1020
QY 1314 cagaatcgaatcctgagtccttactggaatgtaaatgtgttggtgagcgttaaaatgtct 1373
Db 1021 cagaatcgaatcctgagtccttactggaatgtaaatgtgttggtgagcgttaaaatgtct 1080
QY 1374 gttctgtgtccagttgacgactcgaagcctacgtaacatggaacttaattactctctgg 1433
Db 1081 gttctgtgtccagttgacgactcgaagcctacgtaacatggaacttaattactctctgg 1140
QY 1434 ctccagaggtgggtgagagattctaatattcagggtcctatgtatgaaacaaacagcgttg 1493
Db 1141 ctccagaggtgggtgagagattctaatattcagggtcctatgtatgaaacaaacagcgttg 1200
QY 1494 agatctcttaaaaaaagacaagaaacagagtgagggtggaacttcaaatagagtagac 1553
Db 1201 agatctcttaaaaaaagacaagaaacagagtgagggtggaacttcaaatagagtagac 1260
QY 1554 tgtccgccaagaggaagaaatcttcaataatcttctgcaagcgcattgattaatgtct 1613
```

```
||||| 1261 tgtccgccaagaggaagagaaatcttcaataatcttctcagcgcattgattaatgtct 1320
QY 1614 gcaaaaggaatcgtctcttctgaggtttcttctgcactagagattttctgttttgaaaa 1673
Db 1321 gcaaaaggaatcgtctcttctgaggtttcttctgcactagagattttctgttttgaaaa 1380
QY 1674 tagttcgtctctcgtctgtttttgtttattgaactgtttcatgttttttaaaagacatttg 1733
Db 1381 tagttcgtctctcgtctgtttttgtttattgaactgtttcatgttttttaaaagacatttg 1440
QY 1734 aggggagggaggattattatgaatbggggaaaaaagatttgggttgagactaaaattactcat 1793
Db 1441 aggggagggaggattattatgaatgggaaaaaagatttgggttgagactaaaattactcat 1500
QY 1794 cgtcaaaaataatgtaaaaatagttttggggatccacactatattttgtttgatttttaa 1853
Db 1501 cgtcaaaaataatgtaaaaatagttttggggatccacactatattttgtttgatttttaa 1560
QY 1854 cctttcaacattttcccaatgatttgcagagataaactgcacaattttgcataatcatgat 1913
Db 1561 cctttcaacattttcccaatgatttgcagagataaactgcacaattttgcataatcatgat 1620
QY 1914 actggttcttactccaccagttttctataactactaacaagatggtctctcctagcaaga 1973
Db 1621 actggttcttactccaccagttttctataactactaacaagatggtctctcctagcaaga 1680
QY 1974 ttatgtgtttaatgcttctgtttgggttaaaaaataaaagtcagaaaaaggtggaagtc 2029
Db 1681 ttatgtgtttaatgcttctgtttgggttaaaaaataaaagtcagaaaaaggtggaagtc 1736
```

RESULT 4

AAS18576
ID AAS18576 standard; cDNA; 1517 BP.

XX AAS18576;

DT 12-MAR-2002 (first entry)

DE cDNA encoding human translocating chain-associated membrane protein.

XX Human; translocating chain-associated membrane protein; Biotram; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 153..790

FT /*tag= a

FT /product= "Translocating chain-associated membrane

FT protein, Biotram"

XX CN1310184-A.

XX 29-AUG-2001.

XX 24-FEB-2000; 2000CN-0111729.

XX 24-FEB-2000; 2000CN-0111729.

XX (SHAN-) SHANGHAI SHENGYUAN GENE DEV CO LTD.

PI Mao Y, Xie Y;

XX WPI; 2002-034947/05.

DR P-PSDB; AAU10977.

XX New human transposition chain related membrane protein and its coding sequence -

XX Claim 6; Page 18; 22pp; Chinese.

XX The invention relates to a novel human translocating chain associating

XX 01-MAR-2001.
PD
XX
XX 17-AUG-2000; 2000WO-JP05488.
PF
XX
XX 20-AUG-1999; 99JP-0234764.
PR
XX
XX (SUMU) SUMITOMO PHARM CO LTD.
PA
XX
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
PI
XX WPI; 2001-202940/20.
DR
XX P-PSDB; AAB70695.
DR
XX
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
PT
XX
PS Claim 2; Page 62-65; 79pp; Japanese.
CC
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC encodes a specifically claimed human WAR-1 protein from the present
CC invention.
XX
XX Sequence 1110 BP; 296 A; 241 C; 236 G; 337 T; 0 other;
SQ

Query Match 48.5%; Score 1110; DB 22; Length 1110;
Best Local Similarity 100.0%; Pred. No. 6.4e-282;
Matches 1110; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 477 atggggctccgtaagaagacagaccacgaaccccccggtcttcagcaggaattcatctg 536
DB 1 atggggctccgtaagaagacagaccacgaaccccccggtcttcagcaggaattcatctg 60

QY 537 cagaatcatcgggacatcgctctctcgctggggatgttctctctgctgggctgtgttc 596
DB 61 cagaatcatcgggacatcgctctctcgctggggatgttctctctgctgggctgtgttc 120

QY 597 gaggaacacagcagaagcattccatcggtttctcactcttcagcagtggtgctgtccct 656
DB 121 gagggaaacagcagaagcattccatcggtttctcactcttcagcagtggtgctgtccct 180

QY 657 gcagcagaggaaacagccacgggctcaagtcctctattattatgtgtgcaagatttg 716
DB 181 gcagcagaggaaacagccacgggctcaagtcctctattattatgtgtgcaagatttg 240

QY 717 gccacgggtttcttcacatcgctgggcaatcatttcttcacatcgcacaaatcagaatat 776
DB 241 gccacgggtttcttcacatcgctgggcaatcatttcttcacatcgcacaaatcagaatat 300

QY 777 gtgttggtataaaattacaagagaatcagttccaccacgaacgaacaaacaaagttaac 836
DB 301 gtgttggtataaaattacaagagaatcagttccaccacgaacgaacaaacaaagttaac 360

QY 837 gagtcgtcaggttagtggttctactcttttttctgtatttggggcacattcattta 896
DB 361 gagtcgtcaggttagtggttctactcttttttctgtatttggggcacattcattta 420

QY 897 atctctgaaactgctgtcagaccccaactcttatatggaagggtctgcccatagatg 956
DB 1 atctctgaaactgctgtcagaccccaactcttatatggaagggtctgcccatagatg 956

Db 421 atctctgaaaactgcctgtctcagaccccaactcttatatggaagggtctgcccatagatg 480
QY 957 atgacatttcaaatgaagttttttctacatatcccgagttgggttactgttttcactgtttt 1016
Db 481 atgacatttcaaatgaagttttttctacatatcccgagttgggttactgttttcactgtttt 540
QY 1017 cctgaactctacttccagaaaaaccccaaaaaacagacatccctctgcactgtctacatt 1076
Db 541 cctgaactctacttccagaaaaaccccaaaaaacagacatccctctgcactgtctacatt 600
QY 1077 ggtttcacctctccacattactgagagcttatctctgttacttgaaatcatttgggactt 1136
Db 601 ggtttcacctctccacattactgagagcttatctctgttacttgaaatcatttgggactt 660
QY 1137 ctcttttgggtactgacatttttgtgaattactttccacatgtgcgctgtttttac 1196
Db 661 ctcttttgggtactgacatttttgtgaattactttccacatgtgcgctgtttttac 720
QY 1197 tttagtgatgaaaagtaccagaaaggcatctctctgtgggcatgtgttttctcttgggt 1256
Db 721 tttagtgatgaaaagtaccagaaaggcatctctctgtgggcatgtgttttctcttgggt 780
QY 1257 agacttgtaactttaattgttttcgtaactcactgttgggtttcactgtgggtatcgag 1316
Db 781 agacttgtaactttaattgttttcgtaactcactgttgggtttcactgtgggtatcgag 840
QY 1317 aatcggaactcgtgatcccttacttggaatgtaaatgtgttggaagctaaaattgtctgt 1376
Db 841 aatcggaactcgtgatcccttacttggaatgtaaatgtgttggaagctaaaattgtctgt 900
QY 1377 ctgtcgtccagttgcagatcccaagcctcacgtaacatggaacttaattactctctgtgtt 1436
Db 901 ctgtcgtccagttgcagatcccaagcctcacgtaacatggaacttaattactctctgtgtt 960
QY 1437 cagaggtgggttagaagattcttaattattcaggtccctcattgtatgaaaaaagacggtcgaga 1496
Db 961 cagaggtgggttagaagattcttaattattcaggtccctcattgtatgaaaaaagacggtcgaga 1020

QY 1497 tcttctaaaaaaagaaacagaaacggagtgaggagtggaaccttcaaatagatagactgt 1556
Db 1021 tcttctaaaaaaagaaacagaaacggagtgaggagtggaaccttcaaatagatagactgt 1080

QY 1557 ccgcacaaagagaaagagaaatcttctataa 1586
Db 1081 ccgcacaaagagaaagagaaatcttctataa 1110

RESULT 7
AAC77810
ID AAC77810 standard; cDNA; 2833 BP.
XX
AC AAC77810;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:204.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnarary; immunomodulator;
KW antidiabetic; antidiabetic; antirheumatic; antithrombotic; antiviral;
KW antiinflammatory; antihypertensive; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
OS Homo sapiens.
XX
XX WO200055350-A1.
PN
XX 21-SEP-2000.
PD

QY 1940 cataatactaaacagatggtctctctctagcaagattatgtgtttaatgcttgcgttt-ggg 1998
Db 1630 taaaa---tagcagagtgcctgtctctaagaagccataatttttaagtattctttcag 1686
QY 1999 gtaaaataaagtaacaaagagtggaagtcacaa-----tcagttattctgtg 2045
Db 1687 gtaacatggaatactataaagtgtgagtgtaaacctttaatatgttttcagtggtctcta 1746
QY 2046 attgttagaattattttttaagaacttacaaactcagaaaaagattgctagactcaccaaa 2105
Db 1747 atttttttgaatt-ttttagactttcaccttgaaaaaagattgttaaatcacccgga 1805
QY 2106 ataataaatgtcttttattttacagagttagtgattatttagtcttcccatattt--aaa 2163
Db 1806 acaat-tgtgtctttattttatagtgtagtggttatttagttacatccccattttaaa 1864
QY 2164 aaaaacacagttactaagggttaacacatatagaggtttgctgcccataattattgacataaa 2223
Db 1865 acaaaaacataaatggtttacaacacgtggagtttttactaacatacatattataatcaaa 1924
QY 2224 atatcat-taattataataaaattataaaatcattctctgctcatt 2268
Db 1925 gtatatctttaaagtaactgtggaagtaaaatcttcttctgtgcatt 1970
RESULT 8
AAS62269
ID AAS62269 standard; cDNA; 2756 BP.
XX
AC AAS62269;
XX
DT 14-FEB-2002 (first entry)
XX
DE cDNA sequence #56 encoding novel human secreted protein.
XX
KW Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KW immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; gene therapy; antimicrobial; hepatotropic;
KW immunosuppressive; antirheumatic; ss.
XX
OS Homo sapiens.
XX
PN WO200177291-A2.
XX
XX
PD 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US10485.
XX
XX 06-APR-2000; 2000US-195604P.
XX
XX (GEMY) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
XX WPI; 2002-010900/01.
DR
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease -
XX
XX
PS Claim 1; Page 103-104; 391pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple

CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 2756 BP; 825 A; 498 C; 505 G; 928 T; 0 other;

Query Match 40.4%; Score 923.4; DB 24; Length 2756;
Best Local Similarity 75.4%; Pred. No. 1.5e-232;
Matches 1392; Conservative 0; Mismatches 391; Indels 64; Gaps 17;
QY 465 ggaagcgccacatggggctcctgaagaagagcaccagaagcccccggtcttcagccag 524
Db 132 ggcggcccccacatggcgattcgcaagaaagcaccagagcccccgagtgcagccac 191
QY 525 gaattcatctgcagaatcatcgcgacatcgctcctcgctgggagatgtttcttctgctg 584
Db 192 gaattcgctgcagaatcacgcggacatcgctcctcgctgggagatgtttcttctgctg 251
QY 585 gggcttggctcgagggaaacagcagagacatccatcggtttctcactcttcagcacagt 644
Db 252 gggctcatgtttgagataacggcaaaagctctctacattttgttactcttcagtaaat 311
QY 645 gttgctgctcctgcagcagaggaacacgcccgggctcacaagtcctctctattatttgt 704
Db 312 gtaacctcccgacacacagaagaacagtaactgaatcagtgctcctttattactatgc 371
QY 705 gtcaagattggccacgggtttcttctacatgctggtggcaatcattatcattgcacaa 764
Db 372 atcaagattggctactgtttcttctacatgctggtggcaatcattatcattgcgta 431
QY 765 attcagaatattgttgataaaataaagaagaatgcagttcaccaagcgaacaa 824
Db 432 attcaagagtatagttgataaaataaagaagcgaatgcactctcccaacaaacac 491
QY 825 aacaagtttaacgagctggtcagtttagtgttctactttttttcttctgatttggggc 884
Db 492 agcaagtttaataatcgtgctcagtttagtgcgttctacccttttgcgtggttggggc 551
QY 885 acattcattttaatctctgaaaaactgctgcagaccacaaactcttataatgaaagctcgt 944
Db 552 acattcattctctctgaaactacatctcagaccacaaactcttataatgaaagctcgt 611
QY 945 ccccatagcatgatgacatttcaaatgaagtgtttctacatataccagttggcttactgg 1004
Db 612 ccccataacctgatgacatttcaaatgaagtgtttctacatatacagactggcttactgg 671
QY 1005 ttcatgcttttctgaaactacttccagaaaccccaaaaaaacagacatccctctgca 1064
Db 672 ctctcatgttttctgaaactacttccagaaaccccaaaaaaacagacatccctctgca 731
QY 1065 ctgtctacattggtcttccactctccacattactgagacttatctctgttacttgaat 1124
Db 732 ctgtctacattggtcttcttactctctccacattgctgagactgttgaacttgaat 791
QY 1125 cattgggaactctctttgtgactgctgattattttgtgaattacttccacatgtgc 1184
Db 792 catctaggactgtcttctggtgctacattatttttgggaatttcttccacattccc 851
QY 1185 ggcctgttttacttttagtgataaagaccagaaagcagatctctgtgggccaattgtg 1244
Db 852 cgctgttttatttagcaatgaaagatcagaaagattttctctgtgggagattctt 911
QY 1245 ttatcttgggtagacttgta-ctttaattgttccgtactcactgttgggtttcacct 1303
Db 912 ttgttttgggaagacttctgaccttaattcttctcagactgactgtgtgtttggggcct 971
QY 1304 ggcgtggatcgagaaatcggaatcctgtacgtctactgtgaaatgttaaattgttggcagc 1363
Db 972 tgcaagacagaaaaatcagaagcgtgattcagtagtggaaacttcaatgtgttagcgt 1031

Db 601 actgtgtgttttggccttgcaagagcagagaaatcagaaagctggatttcagttactgagaaac 660
Qy 1347 gtaaatgtgtggcagctaaaattgcttctgtctgctcaattgcaagctacccaagctac 1406
Db 661 tcaaatgtgttagctgttagaatcgtctgtctgctgctcatttgcgttactcaggcatt 720
Qy 1407 gtaacatggaacttaattactctctgctcagaggtggttagaagattcttaatttcag 1466
Db 721 atgatgtggaagttcatttaatttcagcttcgaaggtgaggaacattctgttttcag 780
Qy 1467 gcctcatgtatgaaagagaacgctc-----gagattctctaaaaaagaaca 1514
Db 781 gcaccagctgtgaagagaacaaacacagtaactaaaggcagatctcttaaaaaaggaaca 840
Qy 1515 gaaacacgagtg---gagtggaactcaaatagatagactgtccgcgaagagga 1571
Db 841 gaaatggtgtgaatggaacttaactcaaatagcagactctcccggaataaaaa 900
Qy 1572 gagaaattctcatac--tttgcagagcattgatttaattctgcgaaggaattctgctct 1630
Db 901 gagaaattctcatacgaattataaactaattgatttaattgctcccaagaattctgctt 960
Qy 1631 ttgaggtttcttgcactagagattttctgttttgaataat--agtctgtctcttc 1688
Db 961 ctactatattctcagcattagagattttctgttcttgaaatacagctgtgctctt 1020
Qy 1689 ggttttgttaactgtttcatgta-ttttttaagacatttgagggagagagatt 1747
Db 1021 gattttgtctattgcaggttctcatgcaattttttaaaggcattgagggagagattt 1080
Qy 1748 attaatgagggaaaaaagattttgtgtgagactaaattac 1789
Db 1081 gctatgaat--gaaaaaatttttagcttagactaagctac 1120

RESULT 13
AAF74781
ID AAF74781 standard; cDNA; 1092 BP.
AC AAF74781;
XX
DT 17-MAY-2001 (first entry)
XX
DE Rat WAR-1 encoding cDNA sequence SEQ ID NO:4.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder; ss.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..1092
FT FT /*tag= a
FT FT /product= "WAR-1"
XX
PN W0200114582-A1.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-JP05488.
XX
PR 20-AUG-1999; 99JP-0234764.
XX
PA (SUMU) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Okuyama H, Inamura M, Ishikawa H, Nemoto K;
XX
XX WPI; 2001-202940/20.
XX
XX P-PSDB; AAB70696.
XX

PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
XX proteins expressed by the cell -
PS Claim 3; Page 65-68; 79pp; Japanese.
XX
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC encodes a specifically claimed rat WAR-1 protein from the present
CC invention.
XX

SQ Sequence 1092 BP; 260 A; 268 C; 288 G; 276 T; 0 other;

Query Match 28.3%; Score 647.8; DB 22; Length 1092;
Best Local Similarity 76.1%; Pred. No. 3.4e-160;
Matches 844; Conservative 0; Mismatches 247; Indels 18; Gaps 3;
Qy 477 atggggctccgtaagaagcagcaccagaaaccccccttctcagccaggaattcctctg 536
Db 1 atggggctccgcaagaagcagcaccagaaacccccctggtgagccaggaattcctggtg 60
Qy 537 cagaatcatcggaacatgctctctgctgctggtggagctgttctctctctggtggttc 596
Db 61 cagaacacacgagatgctctctgctgctggtggagctgttctctctctggtggttc 120
Qy 597 gagggaacagcagagcattcctctctctctctctctctctcagcagctgttctctct 656
Db 121 gagggaacagcagagcattcctctctctctctctctctcagcagctgttctctct 178
Qy 657 gcagcagagggaacagccacgggctcaggctcctctctctctctctctctctctctct 716
Db 179 -cagcgaagggtaccctcgggtcccgagacccttaccattatgggtcaggatctg 237
Qy 717 gccacggtttct 776
Db 238 gccacggtttct 297
Qy 777 gtgttggaataatacaagagaaatgcagttccaccagcgaagcaaaaaaagtttaac 836
Db 298 gtgttagataagctcagcggagagactgcagtcaccaggaagcaaaaaaagtttaac 357
Qy 837 gagtctgctcagtttagtct 896
Db 358 gaggccggggcagctgagtggtctctctctctctctctctctctctctctctctctct 417
Qy 897 atctctgaaactgcctctcagacccactcttatatggaaggctcgtcccccagcagtg 956
Db 418 gcctctgagaactgctctcagacccactctattgtggaagctcagccccacacatg 477
Qy 957 atgacatttcaaatgaaagtcttctctctctctctctctctctctctctctctctct 1016
Db 478 atgacatttcaagtgaaattttctctctctctctctctctctctctctctctctct 537
Qy 1017 cctgaactctactctcagaaaaacacacacacacacacacacacacacacacacacac 1076
Db 538 ccggagctctactctcagaaagtcaggaaacacacacacacacacacacacacacacac 597
Qy 1077 ggtctacactctctcagacattactgagcttatctctctctctctctctctctctctct 1136
Db 598 ggcctccacctctctcagacattgaggggctctctctctctctctctctctctctctct 657

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 15:16:46 ; Search time 97.09 Seconds
(without alignments)
5788.541 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcacgcgtgtcgca.....ccactgttaaatggaattc 2288

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.2	2.5	7218	1	US-08-232-463-14
2	44	1.9	1431	4	US-09-316-083-2
3	41.2	1.8	5506	4	US-09-004-838-93
4	40.8	1.8	9636	1	US-08-323-170B-1
5	40.8	1.8	9636	4	US-08-954-441-1
6	39.8	1.7	549	4	US-09-342-681C-6
7	39.8	1.7	84495	4	US-09-797-906-3
8	39.6	1.7	19124	2	US-08-487-826B-13
9	39.4	1.7	2791	4	US-09-570-367C-1
10	39.2	1.7	8585	1	US-08-030-096-3
11	38.6	1.7	767	4	US-08-998-416-472
12	38.2	1.7	583	4	US-09-123-912-100
13	38	1.7	2301	4	US-09-561-825-1
14	38	1.7	2361	4	US-09-561-825-26
15	38	1.7	2362	4	US-09-561-825-27
16	38	1.7	2363	4	US-09-561-825-28
17	37.6	1.6	1018	1	US-08-444-083-6
18	37.6	1.6	1018	1	US-08-286-304-6
19	37.6	1.6	1018	1	US-08-442-745-6
20	37.6	1.6	1018	1	US-08-443-129-6
21	37.6	1.6	1018	1	US-08-443-952-6
22	37.6	1.6	1018	1	US-08-443-130-6
23	37.6	1.6	1018	3	US-08-898-911-6
24	37.6	1.6	1018	5	PCT-US95-04467-6
25	37	1.6	921	4	US-09-377-648-4
26	36.8	1.6	19124	2	US-08-487-826B-13
27	36.4	1.6	2361	4	US-09-561-825-29

28 36.4 1.6 4403765 4 US-09-103-840A-2
29 36.2 1.6 3750 3 US-08-617-860B-19
30 35.8 1.6 782 4 US-09-007-119-15
31 35.8 1.6 3850 3 US-08-329-799-34
32 35.6 1.6 860 4 US-08-998-416-287
33 35.6 1.6 7218 1 US-08-232-463-14
34 35.4 1.5 6216 3 US-09-213-053-1
35 35.2 1.5 663 4 US-08-998-416-131
36 35.2 1.5 711 4 US-08-998-416-786
37 35.2 1.5 724 4 US-08-998-416-683
38 35.2 1.5 732 4 US-08-998-416-1036
39 35.2 1.5 828 4 US-08-998-416-538
40 35.2 1.5 854 4 US-08-998-416-534
41 35.2 1.5 860 4 US-08-998-416-287
42 35.2 1.5 1736 3 US-09-182-816-22
43 35.2 1.5 1736 3 US-09-182-816-22
44 35.2 1.5 1736 3 US-09-471-528-22
45 35.2 1.5 1736 3 US-09-471-528-24

RESULT 1

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

ALIGNMENTS

Db 4971 ATTAAATTAATAATATGATGTGTAGTACTTAAATGTGTTGGAATAATGTTATTTT 4912
QY 2052 agaatttatttttaagaacttacaactcagaaagattgctagactcaccataata 2111
Db 4911 AYTGTGTAGTGTGAATAAAGATGGAGGCAAGAGAGAGAGAGAAACGGATATA 4852
QY 2112 aatgttctttttttcacaggtagattattagtgcttccatcccccatttaaaaaacaca 2171
Db 4851 AACTGTGTGCACAAATGAGTTAAGTGAGATTGACTAGTTCCTACCACTTTAAAA 4792
QY 2172 gtactaatgggtaacacatagaggttgctgctccatatatttgcatacaaatatcatt 2231
Db 4791 ATCATTTACATATACAAATATACATGACTACAGAAAAATAATTTTATTATATATCGCT 4732
QY 2232 aattaataaaaaattataatcattctcgtccattccacttgcacttgtaaatgggaatt 2287
Db 4731 TTGTAATACAAATAATATAAACTGTTTAATTAACATATACCAATTTCAATTGTAAAT 4676
RESULT 4
US-08-323-170B-1
; Sequence 1, Application US/08323170B
; Patent No. 5733772
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/323,170B
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Quine, Jonathan A.
; REGISTRATION NUMBER: P-41,261
; REFERENCE/DOCKET NUMBER: 015280-113100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-323-170B-1

Query Match 1.8%; Score 40.8; DB 1; Length 9636;
Best Local Similarity 49.3%; Pred. No. 0.83;
Matches 134; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 1856 tttaaacatttcttaattgatttgcagagataaactgcacaaatttgcataatgatac 1915
Db 3000 TTGATACATTACCTTAAATGTAATTAATAAACACAATATCCAAATATCGAGATAT 3059
QY 1916 tgggtcttactccaccagtggttctataactaacaagatggctctctctctagcaagatt 1975
Db 3060 TTCTTAAACATTTAAAGATATAAAGGAAGTATTAAATTAGATCTTGATATACAAATATC 3119
QY 1976 atgtgtttaaagcttctgttggggttaaaataaaaagtcag-aaaaaggtggaagtcaaatc 2034
Db 3120 AAATGTTTAGTAAATTTTAAATTCATACACAGAAATGCAAAAGTATTTAAATTTATATC 3179
QY 2035 agtattctgtaattgttagaatttttttaagaacttacaactcagaaaaagattgcta 2094
Db 3180 CATATTATTAAATTTTCCATTTTAATCATATAGAAAAAAGAAATTAATAAATAATCCCTA 3239
QY 2095 gactcccccataataaataatgttcttattttt 2126
Db 3240 CATATAAAATCATAAAGATGTGAAATATTTT 3271
RESULT 5
US-08-954-441-1
; Sequence 1, Application US/08954441
; Patent No. 6316000
; GENERAL INFORMATION:
; APPLICANT: Williamson, Kim C.
; APPLICANT: Kaslow, David C.
; TITLE OF INVENTION: Cloning and Expression of Plasmodium
; TITLE OF INVENTION: falciiparum Transmission-Blocking Target Antigen, Pfs230
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,441
; FILING DATE: 20-OCT-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/323,170
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/010,409
; FILING DATE: 29-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 015280-113110US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 149..9556
US-08-954-441-1

Query Match	1.8%;	Score 40.8;	DB 4;	Length 9636;
Best Local Similarity	49.3%;	Pred. No. 0.83;		
Matches 134;	Conservative 0;	Mismatches 137;	Indels 1;	Gaps 1;
QY 1956	tttcaacattttcctaattgatttcgagagataaactgcacaatttgcataatcaatgataac	1915		
Db 3000	TTGTATACATTTACCCTTAAATGTGAATGTTAATAAACAACAATATCCAAATATCGAGATAT	3059		
QY 1916	tgggtcttactccaccagtgtttcataaataactaacaagatgggtctctcctagcaagatt	1975		
Db 3060	TTCTTAAMACATTAAAGATAAAAAGGAAGTATTAAAAATTAGTCTTGATATACAAATATC	3119		
QY 1976	atgtgtttaatgcttgcttgggtgaaataaaaagtacg-aaaaaggtgggaagtcacaatc	2034		
Db 3120	AAATGTTTAGTAAATTTTTTAAATTCATATACACAGAATGCAAGTATTTAAATTTTATATC	3179		
QY 2035	agtatcttgtaattgttagaatttattttttaaagaacttacaactcgaaagaatttgcta	2094		
Db 3180	CATATTATTTAATTTTCCATTAAATCATATAGGAAAAAAGAAATTAATAAATAATCTCTA	3239		
QY 2095	gactcacaaataataaaatgttcttttatttt	2126		
Db 3240	CATATAAAATCATAAAGATGCTGAAATATTTTT	3271		

```

RESULT      6
US-09-342-681C-6
; Sequence 6, Application US/09342681C
; Patent No. 6355782
; GENERAL INFORMATION:
; APPLICANT: Zonana et al.
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins
; FILE REFERENCE: 52978
; CURRENT APPLICATION NUMBER: US/09/342,681C
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 60/092,279
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 60/112,366
; PRIOR FILING DATE: 1998-12-15
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-342-681C-6

```

Query Match	1.7%	Score 39.8;	DB 4;	Length 549;
Best Local Similarity	47.1%;	Pred. No. 0.35;		
Matches 122;	Conservative 0;	Mismatches 137;	Indels 0;	Gaps 0;
QY 1840	gttttgattttaacccttccaacat	tttctctaagatttgcagagataa	actgcacaattt	1899
Db 106	ggtttgatagttggtactcttgc	caaaagccctgaccccttgctg	tgagactccctcaaat	165
QY 1900	tgcatacaatgataactggttct	tactccccacagtggttctoa	ataactacaagaatggt	1959
Db 166	tgcagtgctctgggagtcctcc	ttagtactattttagaaaata	aaacattttctgtctat	225
QY 1960	ctctccctagcaagattatgtgt	ttaatgctgtcttgggtaaa	ataaaagtacgaaaaa	2019
Db 226	ttccaatgactaatatctattt	ttttttcttatggcccgatt	taaaacacagggaaaaa	285
QY 2020	gggtggaagcaaatcagatatt	ctgtaattgttagaattattt	tttaagaactcacaa	2079
Db 286	gggtgaagtctgactttataaa	attgctgctgtgcataatat	tttctaaagttagaaga	345
QY 2080	cagaaaagattgctagact	2098		
Db 346	aaaaaacaaagatgcgatt	364		

```

RESULT 7
US-09-797-906-3
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: ZIANGHE YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(84495)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

```

	Query Match	1.7%; Score 39.8; DB 4; Length 84495;
	Best Local Similarity 51.4%; Pred. NO. 4.6;	
Matches	92; Conservative 0; Mismatches 87; Indels 0; Gaps 0;	
Qy 1684	tcttcggttttgcattgaactttccatgtattttaagaacatttgaggggagagg 1743 	
Db 7178	ttttcttcttctattattattattattattattcattgttttggggtacaagg 7237 	
Qy 1744	gattattatgaatggaaaaaaagttttggtagacctaatattactcatcgtaaaaata 1803 	
Db 7238	tagttttcgtttacatggatgaattactttaagatgaattctgatatttgtgcacca 7297 	
Qy 1804	atgtcaaaatagttttggggaccacctatattttttgattttttaacctttcaac 1862 	
Db 7298	tcaccaagcaagtgcattgccaggatgtttttcttttatccccttcacccctgcac 7356 	

RESULT 8
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827

```

1  RESULT 8
2  US-08-487-826B-13
3  ; Sequence 13, Application US/08487826B
4  ; Patent No. 5993827
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Sim, Kim L.
7  ; APPLICANT: Chitnis, Chetan
8  ; APPLICANT: Miller, Louis H.
9  ; APPLICANT: Peterson, David S.
10 ; APPLICANT: Su, Xin-zhaun
11 ; APPLICANT: Wellemns, Thomas E.
12 ; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
13 ; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
14 ; NUMBER OF SEQUENCES: 45
15 ; CORRESPONDENCE ADDRESS:
16 ; ADDRESSEE: Knobbe Martens Olson & Bear
17 ; STREET: 620 Newport Center Drive 16th Floor
18 ; CITY: Newport Beach
19 ; STATE: California
20 ; COUNTRY: US
21 ; ZIP: 92660
22 ; COMPUTER READABLE FORM:
23 ; MEDIUM TYPE: Floppy disk
24 ; COMPUTER: IBM PC compatible
25 ; OPERATING SYSTEM: PC-DOS/MS-DOS
26 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
27 ; CURRENT APPLICATION DATA:
28 ; APPLICATION NUMBER: US/08/487,826B
29 ; FILING DATE: 10-SEP-1993
30 ; CLASSIFICATION: 435
31 ; ATTORNEY/AGENT INFORMATION:

```



```

; ORGANISM: Glycine max cv. Williams 82
;
; FEATURE:
;
; NAME/KEY: promoter
; LOCATION: (1)..(2362)
US-09-561-825-27

```

	Query Match	1.7%	Score 38;	DB 4;	Length 2362;	
	Best Local Similarity	50.0%	Prod. No. 2.2;	Mismatches	95; Indels	
	Matches	95;	Conservative	0;	Gaps	
QY	1698	tattgaactgtttc	atgtattttt	taagacattt	gaggggagggaggtatttatgaa	gt 1757
Db	955	TATCGAAATGATTATTCGTAAAGGAAAGAAATCTTTTAAAGTAGTCCTCTATTCTCTAAAT	896			
QY	1758	ggaaaaaagaat	tgttgggttgagactaaat	taactcgttcaaaa	taatgtccaaatagtt	1817
Db	895	AAAAAGTAAAAATGATTTTCGCTTTAAATTAATATTTATCGAACATTTCTATTGAAAA	836			
QY	1818	ttggggatccacac	catattt	gtgtttgtattt	taacctttcaacattttcc	taagatt 1877
Db	835	TTTAAGATAAATATTAATTTTAAAGATAAATGAATTTTCATAAATGATCTATTATA	776			
QY	1878	tgcagagata	1887			
Db	775	TCAAAAAAAA	766			

Search completed: September 6, 2002, 15:18:39
Job time: 7823 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2002, 13:07:54 ; Search time 3580.18 Seconds
(without alignments)
8625.553 Million cell updates/sec

Title: US-09-807-470-3

Perfect score: 2288

Sequence: 1 tataggcgacgggtgtcg.....ccactgttaagggaattc 2288

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_gss:*

13: em_gss_hum:*

14: em_gss_inv:*

15: em_gss_pln:*

16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	785.8	34.3	1356	11 BC020283	BC020283 Homo sapi
2	698.2	30.5	760	10 BG714759	BG714759 602673845
3	692.6	30.3	1064	9 AL523342	AL523342 AL523342
4	608.2	26.6	1142	10 BM466240	BM466240 AGENCOURT
5	577.6	25.2	1135	10 BM466239	BM466239 AGENCOURT
6	555.2	24.3	822	9 AU130188	AU130188 AU130188
7	553.2	24.2	814	9 AU132871	AU132871 AU132871
8	532.2	23.3	779	10 BM045930	BM045930 603624955
9	511.4	22.4	790	10 BE612552	BE612552 601452065
10	509.4	22.3	879	10 BI870787	BI870787 603394607
11	508.2	22.2	905	9 AU131342	AU131342 AU131342
12	503.8	22.0	727	10 BG547017	BG547017 602573813
13	500.6	21.9	836	10 BF969571	BF969571 602271922
14	497.4	21.7	879	9 AU128453	AU128453 AU128453
15	495.2	21.6	910	10 BM459490	BM459490 AGENCOURT
16	483.6	21.1	944	10 BI655670	BI655670 603283957
17	483	21.1	483	9 AV725026	AV725026 AV725026

18	481.8	21.1	769	10 BG542329	BG542329 602571885
19	481.8	21.1	808	10 BI830439	BI830439 603073495
20	476	20.8	496	9 AW078666	AW078666 xb02907.x
21	472.2	20.6	660	10 BF984410	BF984410 602307936
22	468	20.5	969	10 BG975574	BG975574 602845246
23	462.8	20.2	1015	10 BG389557	BG389557 60214256
24	462	20.2	878	10 BE618241	BE618241 601462496
25	459	20.1	695	10 BI733351	BI733351 603353253
26	453.6	19.8	651	10 BG714581	BG714581 602671468
27	453	19.8	476	9 AI138349	AI138349 qd87a01.x
28	451.8	19.7	859	10 BG323904	BG323904 602422933
29	447.2	19.5	636	10 BG077352	BG077352 H3014E03-
30	445	19.4	787	10 BE788046	BE788046 601482653
31	444	19.4	450	9 AW023347	AW023347 df53f06.y
32	444	19.4	608	10 BG614318	BG614318 602642420
33	444	19.4	614	9 AL602739	AL602739 DKZP6866F
34	442.8	19.4	1067	10 BF538214	BF538214 602053686
35	436.8	19.1	944	10 BG916945	BG916945 602816230
36	433.6	19.0	504	12 AQ127602	AQ127602 HS_3061_A
37	432.6	18.9	493	9 AA169554	AA169554 zp17b09.s
38	432	18.9	835	10 BE887450	BE887450 601509592
39	426.6	18.6	569	9 AT201635	AT201635 qb81d09.x
40	424.2	18.5	825	10 BI731069	BI731069 603351703
41	423.8	18.5	736	10 BF141338	BF141338 601789863
42	422	18.4	1048	9 AL523341	AL523341 AL523341
43	421.4	18.4	708	10 BE914578	BE914578 601665679
44	418.4	18.3	773	10 BI730813	BI730813 603351811
45	417.6	18.3	731	10 BG700572	BG700572 602680806

ALIGNMENTS

RESULT 1

BC020283	BC020283	Homo sapiens, clone IMAGE:4519858, mRNA.	1356 bp	mRNA	linear	HTC 20-DEC-2001
LOCUS	BC020283					
DEFINITION	BC020283					
ACCESSION	BC020283.1	GI:17946782				
VERSION	BC020283.1					
KEYWORDS	HTC.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 1356)					
AUTHORS	Strausberg,R.					
TITLE	Direct Submission					
JOURNAL	Submitted (20-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA					
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov					
COMMENT	Contact: MGC help desk Email: cgaps-r@mail.nih.gov Tissue Procurement: DCTD/btp cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: B#CM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: villalon@bcm.tmc.edu Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gunaratne, P., Yoon, V., Kowis, C., Martin, R., Lawrence, S., Richards, S., Gibbs, R.A. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAC Plate: 29 Row: h Column: 1 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis					


```
RESULT 5
LOCUS BM466239 1135 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6457038 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5576991
5', mRNA sequence.
ACCESSION BM466239
VERSION BM466239.1 GI:18515281
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1135)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12329 row: 1 column: 16
High quality sequence stop: 710.
FEATURES
source
location/Qualifiers
1..1135
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5576991"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
1 others
BASE COUNT 285 a 259 c 241 g 349 t
ORIGIN
Query Match 25.2%; Score 577.6; DB 10; Length 1135;
Best Local Similarity 82.9%; Pred. No. 9.2e-108;
Matches 695; Conservative 0; Mismatches 139; Indels 4; Gaps 3;
QY 465 ggaagcgccaccatgggctccgtaagaagcagcaccagaaacccccccgttctcagccag 524
DB 41 GCGGCCCCCACCATGCGGATTCGCAAGAAAGCACCAGAGCCCCCAGTGTGAGCCAC 100
QY 525 gaattcctctgcagaatcgtggacatcgctcctcgtggtgggatgttcttctgtgtg 584
DB 101 GAATTCGCTGCTCAGAAATCAGCGGACATCGTCTGCTGTGTGGCATGGTCTTCCTGCTG 160
QY 585 gggctgtgttcgaggaacagcagacatccatcggtttctcactcttcacagcagcagt 644
DB 161 GGGCTCATGTTTGAGATACGGCAAAAGCTTCTATCATTTTTTGTGTACTCTTCAGTCAAT 220
QY 645 gttgtgtcctgcagcagagaaacagccagggctcaaaqtcctctattattatggt 704
DB 221 GTCAACCTCCACGACACAGAGAACAGACTACTGAATCAGTGTCTCTTTATTACTATGGC 280
QY 705 gtcaagaatttgccacaggttttctctacatgctgtggtggcaatcattattccatccaca 764
DB 281 ATCAAGAAGATTTGGCTACTGTTTCTTCTACATGCTAGTGGCGATAATATTATCCGCGTA 340
QY 765 attcaggaatattgtttggataaaattacaagaagaatgcagttccaccagaagcgaacaa 824
DB 341 ATTTCAAGAGTATATTGTTGGATAAAAATTAACAGCGCAATCTCTCCAAACAAACAC 400
QY 825 aacaagttaacagcagtcgtgcagtttagtgtgttctacttttttcttctgtattggggc 884
```

```
Db 401 AGCAAGTTTAAATGAATCTGGTCAGCTTAGTGGCTTCTACTCTTTTGCCTGTGTTGGGC 460
QY 885 acattcattttaaattctctgaaacacgctgtcagagcccaactcttatatggaagctcgt 944
Db 461 ACATTCAATTCATCTCTGAAAACTACATCTCAGAGCCCAACTATCTTATGAGGGCTTAT 520
QY 945 ccccatagcatgatcacatttcaaatgaagtttttctacatcccatcagttggtctactg 1004
Db 521 CCCCAATAACCTGATGACATTTCAATGAAGTTTCTTACATATCACAGCTGGCTTACTGG 580
QY 1005 ttctatgcttttcctggaactctacttccagaaaccccaaaacagacatccctcgtcaa 1064
Db 581 CTTCATGCTTTTCTGAACTCTACTTCCAGAAACCAAAAGAGATATTCTCTGTCAG 640
QY 1065 ctgtctacattggtcttccactcttccacattactgagcttactcttctgtactgaat 1124
Db 641 CTGTGCTACATTTGGCTTTTACCTCTTCCACATTTGCTGGAGCTTACCTTTGAACCTGA 700
QY 1125 catttgggaactctcttcttggtagctgacttatttggtaatttcttccacacatgtgc 1184
Db 701 CATCTAGGACTTGTCTTCTGCTGTCTACATTTTGTGTAATTTCTTTCCACATTTTC 760
QY 1185 ggcctgttttacttagtgat-gaaaagtaccagaaaggca-tatctctgtggcca--t 1240
Db 761 CCCTGTTTTATTATTAGCAATGGAAGATATCAGAAAGGATTTCTCTGTGGCCAGNT 820
QY 1241 tgtgtttatcttggtagactgtgactttaattgttccgtactactcactgttgggttt 1298
Db 821 CTTTGTGTTTTGGGAAAACCTCTGACTTTAATCTCTTTCAGAACTTGACTGGGTGTTT 878

RESULT 6
LOCUS AU130188 822 bp mRNA linear EST 24-OCT-2000
DEFINITION AU130188 NT2RP3 Homo sapiens cDNA clone NT2RP3000388 5', mRNA
sequence.
ACCESSION AU130188
VERSION AU130188.1 GI:10990542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 822)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5' & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
location/Qualifiers
1..822
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3000388"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 213 a 171 c 148 g 287 t
ORIGIN
```

```
Query Match      24.3%; Score 555.2; DB 9; Length 822;
Best Local Similarity 82.0%; Pred. No. 3.4e-103;
Matches 638; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy 622 tggtttcactcttcagcacagtggtgctgcccgcagacagagagacagcccggtct 681
Db 1 TTTTGTGTACTCTTCAGTACAATGTCACCCCTCCGACGACAGAGAACAGCTACTGAAT 60

Qy 682 caaagtcctctattattatgtgtcaaaagattggccacggttttcttcacatctgg 741
Db 61 CAGTGTCCCTTTATCTACTGGCATCAAGGATTTGGCTACTGTTTCTTCTACAGCTAG 120

Qy 742 tggcaatcattatcagcccaaatccaggaaatgtgttgataaaataaacaagagaa 801
Db 121 TGGCGATAATTATTATCGCGCTAATCAAGAGATATATTGGATAAAATTAAACAGCGAA 180

Qy 802 tgcagttccaaagcgaagaaacaaacaaagtttaacagagtcgtgctagttgttct 861
Db 181 TGCACCTTCTCAAAACAAAACACAGCAAGTTTAATGAATCTGGTCAGCTTAGTGCCT 240

Qy 862 actttttctgtatttgggacattctttaaactctctgaaactgctgctcagacc 921
Db 241 ACCTTTGTGCTGTGTTGGGACATTCATCTCATCTCTGAAACATACATCTCAGACC 300

Qy 922 caactcttatatgaaggtcgtcccatagcatgatgacatttcaaatgaagttttct 981
Db 301 CAACATCTTATGAGGGCTTATCCCCATACCTGATGACATTTCAAATGAAGTTTCT 360

Qy 982 acatatccagatccctcgtcaactgtctacattgtgttccttcaactcttccaaattactg 1101
Db 361 ACATATCACAGCTGGCTTACTGGCTTCATGCTTTTCTGAACTCTACTTCCAGAAACCA 420

Qy 1042 aaaaacaagacatccctcgtcaactgtctacattgtgttccttcaactcttccaaattactg 1101
Db 421 AAAAAGAAGATATCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

Qy 1102 gagcttatctgtacttgatcaatttgggaactctcttcttggctactgctgattttg 1161
Db 481 GAGCTTACCTTTTGAACCTTGAATCATCTAGGACCTGCTGCTGCTGCTGCTGCTGCTG 540

Qy 1162 ttgaattacttccacatgtgcgctctgttttacttagtgatgaagaagaccagaaag 1221
Db 541 TTGAATTTCTTTTCCACATTTCCGCGCTGTTTATTTATTTAGCAATGAAAGATATCAGAAAG 600

Qy 1222 gcatatctgtggccatgtgtttatctgttggttagactgtgactttaaatttttccg 1281
Db 601 GATTTTCTGTGGCAGTCTCTTTTGTGTTGGGAAGACTTCTGACTTAAATCTTTTTCAG 660

Qy 1282 tactcaactgttgggtttcacctggtggtgatcgagaaatcggaatcctgatgaccttactg 1341
Db 661 TACTGACTGTGTTTGGCCCTTCGAAGAGCAGAGAAATCAGAACTGGATTTTCAGTACTG 720

Qy 1342 gaaatgtaaatgtgtggcagctaaaattgtctgtctgtccagttgcacgatcca 1399
Db 721 GAAACTTCAATGNGTGTAGCTGCTGAGATCGTGTGCTGGCATNCATTTGCGTTACTCA 778
```

```
RESULT 7
AUI32871      AUI32871      814 bp      mRNA      linear      EST 24-OCT-2000
LOCUS        AUI32871 NT2RP4 Homo sapiens cDNA clone NT2RP4000757 5', mRNA
DEFINITION
sequence.
ACCESSION    AUI32871
VERSION      AUI32871.1 GI:10993410
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 814)
AUTHORS      Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,
```

Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and Isogai,T.
HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y., Isogai,T.)
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES
Location/Qualifiers
1..814
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4000757"
/clone_lib="NT2RP4"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/notes="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 206 a 195 c 161 g 249 t 3 others
ORIGIN

Query Match 24.2%; Score 553.2; DB 9; Length 814;
Best Local Similarity 82.4%; Pred. No. 8.8e-103;
Matches 645; Conservative 0; Mismatches 136; Indels 2; Gaps 1;

```
Qy 465 ggaagcgccaccatggggctccgttaagaagagcaccagaaccccccgcttctcagcag 524
Db 33 GCGCGCCCCACCATGGCGGATTTCGCAAGAAAAGCACCAGAGCCCCCAGTGTGAGCCAC 92

Qy 525 gaatttcactctcagaatcatcgagacatgctctccctgcgtgggagtgcttctcctctg 584
Db 93 GAATTCGTCTGCGAAGATACGCGGACATGCTCTCTGTGTGGCGATGGTCTTCTCTGTG 152

Qy 585 gggctgtgttcagggaacagcagaagcatccatcgtgtttctcactcttcacacagct 644
Db 153 GGGCTCATGTTTGAGATAACGGCAAAAGCTTCTATCATTTTGTGTACTTCTTCAGTACAAT 212

Qy 645 gttgctgcccctcagcagaggaacagcagccagggctcgaagtcctctctattatgggt 704
Db 213 GTCACCCCTCCAGCAACAGAGAACAAAGCTACTGAATCAGTGTCCCTTTATTTACTATGGC 272

Qy 705 gtcaaaagatttggccacggttttctctacatgctgtggaatcattattcatgcccaca 764
Db 273 ATCAAAAGATTGGCTACTGTTTCTTCTACATGCTAGTGGGATTAATATTTCATGCCGTA 332

Qy 765 attcaggaatatgtgttggataaaataaacaagagaatgcagttccaccaagcgaaacaa 824
Db 333 ATTCAAGAGTATATGTTGGATAAAATTAACAGGCGAATGCACCTTCTCCAAAACAAACAC 392

Qy 825 acaagtttaacagagtcgtgctcagtttagtgttcttcaacttttttcttatttggggc 884
Db 393 AGCAAGTTTAATGAATCTGCTAGCTTAGTGGCTTCTACCTTTTGGCTGTGTTGGGGC 452

Qy 885 acatttcatttaactctctgaaactgcctgcagacccaactcttatatgaaggtcgtg 944
Db 453 ACATTCAATCTCATCTCTGAAACTACATCTCAGACCCCACTATCTTATGGAGGCTTAT 512

Qy 945 ccccatagcatgatgacatttcaaatgaagtgttttctacatatccccagttggcttactgg 1004
Db 513 CCCCAATCACTGATGACATTTCAAATCAAGTTTCTTACATATACACAGCTGGCTTACTGG 572

Qy 1005 ttcatgcttttctcgtgaactctacttccagaaaaaaccaaacagacatccctcgtcaa 1064
```



```

Db 417 CCGTAATCAAGAGTATAGTTGGATAAAATTAACAGGCGAATGCACCTTCTCCAAACAA 476
QY 820 aacaaacaagtttaacagagctgtgctagcttagtgctgtctactttttttctgtattt 879
Db 477 AACACAGCAAGTTAATGAATCTGCTAGCTTAGTGCCTTCTACCTTTTTCCTGTGTTT 536
QY 880 ggaggacatttatttaactctgaaactgcctgcagacccaactcttatatgaagg 939
Db 537 GGGGCACATTTCATCTCTGAAACTACATCTCAGACCCCACTATCTTATGAGGG 596
QY 940 ctctgtcccatcagcatgatgacatttcaaatgaagtctttctacatatccagttggctt 999
Db 597 CTTATCCCCCAATCAACCTGATGACATTTCAATGAAGTTTTCCTACATATCACAGCTGGCTT 656
QY 1000 actgtgttcaatctttctcgaactctac-ttccagaaacaaacaaacaaacaaacacccct 1058
Db 657 ACTGGCTTCATCTGCTTTCCTGAACTCTACTTTCCAGAAACCAAAAGAGATATTCCT 716
QY 1059 cgtcaactgtctacatgtgtctccactctccacattcctcagagcttatctctgtac 1118
Db 717 CGTACCTGTCTACATGGTCTTTACCTCTCCACATTCGTGGAGCTTACCTTTTGAAAC 776
QY 1119 ttgaatactttgggactcttcttctgtgactgcatattttgttgaattactttccac 1178
Db 777 TTGAATCATCTAGGACTTCTCCTGCTGCTACATTAATTTTGTGGATTCTTTTCCGG 836
QY 1179 atgtggcgcctttttacttttag 1201
Db 837 CATTTCCCGCTGGTTAATTTTAG 859

RESULT 11
AUI31342
LOCUS AUI31342 NT2RP3 Homo sapiens cDNA clone NT2RP3002413 5', mRNA
DEFINITION
ACCESSION AUI31342
VERSION AUI31342.1 GI:10991696
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 905)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5', & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers
1..905
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3002413"
/clone_lib="NT2RP3"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/note="Vector: pME18SFLJ; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 214 a 221 c 212 g 255 t
ORIGIN
```

```

Query Match 22.2%; Score 508.2; DB 9; Length 905;
Best Local Similarity 81.8%; Pred. No. 1.4e-93;
Matches 597; Conservative 0; Mismatches 131; Indels 2; Gaps 1;

QY 465 ggaagcgccaccatgggctccgtaagaagagcaccacaagcccccgcttctcagccag 524
Db 176 GCGGCCCCACCACATGGCATTCGCAAGAAAAGCACAAGAGCCCCCAGCTGTGAGCCAC 235
QY 525 gaattactctcgcagaatcgcgcacatcgtctcctcgtcgtgggagtgcttctctcgtg 584
Db 236 GAATTCGCTCTCAGAAATCACGCGGACATCGTCTCTGTGTGGCGATGCTCTCTCTGCTG 295
QY 585 gggcttggttcgaagggaacagcagaatccatccatcggtgtttctcactcttcagcacagt 644
Db 296 GGGCTCATGTTTGAGATAACGGCAAAAGCTTCTATCATTTTGTGTACTCTTCTCAGTACAA 355
QY 645 gtgtgtctccctgcagcagaggaacacacgagcgggtccaaagtcctctctattatttgt 704
Db 356 GTCACCCCTCCAGCAACAGAGAAACAAGCTACTGAATCAGTCTCCCTTTTACTATGGC 415
QY 705 gtcaagatttggccacggtttctctctacatcgtcgttggcgaatcattatcattccaca 764
Db 416 ATCAAGAGATTGGCTACTGTTTCTTCTACATGCTAGTGGCGATAATATTATCATGCCGTA 475
QY 765 attcaggaatatgttgataaaattacaagagaatgcaggtccacaaagcgaacaa 824
Db 476 APTCAAGAGTATATGTTGATAAAATTAAACAGGCGAATGCACCTTCTCCAAAACAAACAC 535
QY 825 aacaagtttaacgagtcctggtcagtttagtgtgtctacttttttctgtatttggggc 884
Db 536 AGCAAGTTTAATGAATCTGGTCAGCTTAGTGGCTTCTACCTTTTGGCTGTCTTTGGGC 595
QY 885 acattcatttaactctcgaagcgcctgctcagacccaactcttatatggaagctcgt 944
Db 596 ACATTCAATTCATCTCTGAAACATACATCTCAGACCCCAACATATCTTTGAGGGGCTTAT 655
QY 945 ccccatagcatgacacatttcaaatgaagttttctacatcccatccagttggcttactgg 1004
Db 656 CCCCATACCTGATGACATTTCAAATGAAGTTTCTCATATACAGCTGGCTTACTGG 715
QY 1005 ttctcgtctttccctgaactctactctccagaaacacaaacaaacagacatccctcgtcaa 1064
Db 716 CTTTCATGCTTTTCCTGAACCTCTACTTNCAGAAACCAAAAGAGATATCTCTGCTCAG 775
QY 1065 ctgtctacatggttcttccactctccacatctcagagcttatctctgtacttgaa 1124
Db 776 CTGGCTACATTGGTCTTTACCTCTTTTACATTTGCTGGAGCTTACCTTTTGAACCTGAAT 835
QY 1125 catttgggactctctcttttggtaactgaattatttggtaattacttccacacatgctg 1184
Db 836 CATCTAGGACTGGTCTCTGGG--CTACATTAATTTGGTGAATTTCTTTTCACATTTNCC 893
QY 1185 ggcctgtttt 1194
Db 894 GCTGTTTATT 903

RESULT 12
BG547017
LOCUS BG547017 727 bp mRNA linear EST 04-APR-2001
DEFINITION 602573813F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702154 5',
mRNA sequence.
ACCESSION BG547017
VERSION BG547017.1 GI:13545682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 727)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
```


TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cygaps-f@mail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CMI538 row: 1 column: 03
 High quality sequence stop: 714.
 Location/Qualifiers
 i. 127
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4702154"
 /clone_lib="NIH_MGC_77"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site_1:
 SfiI (ggcgctctggcc); Site_2: SfiI (ggccattatggc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATATGGC-3', and 3' adaptor sequence:
 5'-ATWCTAGAGCCGAGCGCGGCATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size 1.9
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
 by PCR. This library was enriched for full-length clones
 and was constructed by Clontech Laboratories (Palo Alto,
 CA). Note: this is a NIH_MGC Library."
 190 a 176 c 128 g 233 t
 BASE COUNT 190 a 176 c 128 g 233 t

BASE COUNT	190 a	176 c	128 g	233 t	CA). Note: this is a NIN_MGC Library.
ORIGIN					
Query Match	22.0%	Score 503.8	DB 10	Length 727	
Best Local Similarity	84.0%	Pred. No. 1.1e-92			
Matches 604	Conservative 0	Mismatches 112	Indels 3	Gaps 3	
QY	486	cgtaagaagcagcaccagaaccccccgctctcagccaggaattcctcctgcagaatcat	545		
Db	2	CGCAAGAAAGACCAAGAGCCGCCAGTGTGAGCCACGAATTCGTCTGCAGAAATCAC	61		
QY	546	gcggacatgctctcctgcgtgggagtcttctcgtctggggctgtgttcagagggaaaca	605		
Db	62	CGGACATCGTCTCCTGTGGCGCATGGCTCTCTGTGGGGCTCATGTTTGAATACAG	121		
QY	606	gcagaagcatccatcgtgtttctcaactctcagcacagtgctgtcctcgcagcagag	665		
Db	122	GCAAAAGCTCTATCATTTTGTGTACTCTTCAGTACAATGTGTCACCTCCCGACAGAA	181		
QY	566	gaacaagccacgggctcaagtcctctattattatgtgtcctaagatttggccacggtt	725		
Db	182	GAACAAGCTACTGAATCAGTGTCCCTTTATTACTATGGCATCAAGAATTTGGCTACTGTT	241		
QY	726	ttctctacatgctgggtggcaatcattatcatgccacaattcaggaatatgtgttgat	785		
Db	242	TTCTTCTACATGTAGTGGCGCAATTAATTCATGCCCGTAATTCGAAGAGTATATGTTGGAT	301		
QY	786	aaataacaagaaatgcagttccacaaagcgaaacaaacaaagtttaacgagctctggt	845		
Db	302	AAATTTAACAGCGGAATGCACCTTCTCCAAAACAAAACACAGCAAGTTTAATGAATCTGGT	361		
QY	846	cagtttaagtgttctacttttttctgtattttgggggcacattcatttaactctcga	905		
Db	362	CAGCTTAGTCGTCTACCTTTTTCCTGTGTGTTGGGGCACATTCATTCATCTCTGAA	421		
QY	906	aactgcctgtcagaccacaactcttatggaagctcgtccccatagcatgacattt	965		
Db	422	AACCTACATCTCAGACCCCAACTATCTTATGGAGGCGTTATCCCCATAACCTGATGACATT	481		
QY	966	caaatgaagtttttctacatatccacagttggcttactgtgttctatgtctttctcgaact	1025		
Db	482	CAATGAAGTTTTTCTACATATCACAGCTGGCTTACTGGGTTTCTATGCTTTTCTCTGAAC	541		

Qy	1026	tacttcagaaaaaccccccaagaacatccctcggtcaacttgtcta-catgtggtttcca	1084
Db	542	TACTTCCAGAAAAACCAAAAGAATATTCTCGTCAGCTTGTTCTACCAATGGTCTTTA	601
Qy	1085	cctttcacattactgagccttacc-tcttgtaactgaatacatctgggacctctcttt	1143
Db	602	CCTCTCCACAATTCGTGAGCTTACCCTTTTGAACATCATACTAGGACTTGTTCTTC	661
Qy	1144	tgttact-gcaattatttggtaattactctcccacatgtgcgcctgttttacttag	1201
Db	662	TGGTGCTAACATTAATTTCTTGTGAATTTCTTTTCCACATTTCCCGCCTGTTTTATCTAAG	720
RESULT 13			
BF969571 836 bp mRNA linear EST 22-JAN-2001			
LOCUS 602271922F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:436085 5',			
DEFINITION mRNA sequence.			
ACCESSION BF969571			
VERSION BF969571.1 GI:12336786			
KEYWORDS EST.			
SOURCE human.			
ORGANISM Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE 1 (bases 1 to 836)			
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.			
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL Unpublished (1999)			
COMMENT Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LHAM10000 row: 1 column: 06 High quality sequence stop: 672.			
FEATURES			
source Location/Qualifiers			
1..836			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:436085"			
/clone_lib="NIH_MGC_84"			
/tissue_type="adrenal cortex carcinoma, cell line"			
/lab_host="DH10B (phage-resistant)"			
/note="Organ: adrenal gland; vector: pCMV-SPORT6; Site:1; Noti; Site:2; SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.229 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."			
BASE COUNT 197 a 211 c 176 g 252 t			
ORIGIN			
Query Match 21.9%; Score 500.6; DB 10; Length 836;			
Best Local Similarity 80.7%; Pred. No. 5e-92;			
Matches 598; Conservative 0; Mismatches 139; Indels 4; Gaps 1;			
Qy	465	ggaagcgccacatggggctcgtaaagagagaccacaagaccccccggtcttcaggcag	524
Db	29	GCGCGCCCCCATYGGCGATTTCGGAAGAAAAGCACCAAGAGCCCCCAGTCTGAGCCAC	88
Qy	525	gaattcacctgcagaatacatgcggacatcgtctcctcgctggggagtcttcttcctgct	584
Db	89	GAATTCGTCCTGCAGATACACGCCGACATCGTCTCCGTGTGGCGATGGTCTTCTGCTG	148
Qy	595	gggctgtgttcgagggaacagcagaagcatccatcgtgtttctcaactcttcagcacagt	644
Db	149	GGGCTCATGTTTGATGAACCGGAAAAGCTCTCTATCAATTTTGTACTTCTTCAGTACAAT	208

BASE COUNT		247	a	169	c	150	g	300	t	13	others	
ORIGIN												
<div> <div> /organism="Homo sapiens" /db_xref="taxon:9606" /clone="NT2RP2003491" /clone_lib="NT2RP2" /cell_type="teratocarcinoma" /cell_line="NT2" /note="Vector: pWE18SFU3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction" cells after 2-weeks retinoic acid (RA) induction" 150 g 300 t 13 others </div> <div> Query Match 21.7%; Score 497.4; DB 9; Length 879; Best Local Similarity 77.7%; Pred. No. 2:2e-91; Matches 620; Conservative 0; Mismatches 175; Indels 3; Gaps 2; </div> </div>												
Qy	711	gatttggccacgggtttctttacatcgtggtgcgaatcattattcatgcacaaattcag 770										
Db	1	GATTTGGCTACTGTTTCTTCTACATGCTAGTGGCATNATTATTATCGCCGTAATCAA 60										
Qy	771	gaatatgttggataaaaattacaagaagaatgcagttccacaaagcgaaacaaaaaag 830										
Db	61	GAGTATATGTTGGATAAAATTAACAGCGGAATGACATCTCCAAACAAAAACAGCAAG 120										
Qy	831	ttaacagctcgtcagtttagtgtctcaacttttttcttctgtatttggggcacattc 890										
Db	121	TTTAATGAATCGTCAGCTTAGTCGGTCTACCTTTTGGCTGTGTTGGGGCACATTC 180										
Qy	891	atttaattctgtaaaactgcgtgcagacccaactcttatatgaaaggtcgtcccaat 950										
Db	181	ATTCTCATCTCTGAAACTACATCTCAGACCAACTATCTTATGGAGGGCTTATCCCAT 240										
Qy	951	agcattgacatttcaaaagaattttctacatatcccaattgggttactgttttcat 1010										
Db	241	AACCTGATGACATTTCAATGAAGTTTTTTCACATATACAGCTGGGTACTGGCTTCAT 300										
Qy	1011	gcttttccgtgaactctacttcagaaaaacaaaaacaaagacatccctcgtaactgtc 1070										
Db	301	GCTTTTCTCGAATCTACTTCCAGAAAAACCAAAAGAAGATATCTCTCGTCAGCTGTC 360										
Qy	1071	taeattggtttacactcttccacattactggagcttatctcttgtactgaatcattg 1130										
Db	361	TACATTGGCTCTTACCTCTTCCACATTGCTGGAGCTTACCTTTTGAATCAATCTA 420										
Qy	1131	ggactctctcttggctactgcattatttgtgaattactttccacatgtcgcgctg 1190										
Db	421	GGACTTGTCTCTGGTGCTACATATTGTTGCAATTTCTTTTCCACATTTCCGCGCTG 480										
Qy	1191	tttacttttagtgaagaaagtaacagaagagcatactctgtgggccaattgtgtttatc 1250										
Db	481	TTTTATTTTAGCAATGAAAGTAGTATCAGAAAGGATTTCTCTGTGGCAGTGTCTTTTGT 540										
Qy	1251	ttgggtagacttgcatttaattgtttccgtactcaactgttgggtttccactgtcgtgga 1310										
Db	541	TTGGGAANANFTCGACTTAATTTCTTTCAATCTGACTGTGTGTTTGGCTTGCANANA 600										
Qy	1311	tcgcagaatcggaatcgtgatgccttacttggaagtgaattgttggcagctaaatt 1370										
Db	601	ACANAAATCAGAAACTGGATTTTCAGTACTTGAAACTTCAATGTGTTT-ACTGTTAAATC 659										
Qy	1371	gctgtttctgtgtccagtttgcacagatccaaagctcacgtaacatggaacttaattactctc 1430										
Db	660	GCTGTTCTGGGATCCATTTGCGTTTNCACAGGCATTTAATGTGGAAATTTTCATTAATTTT 719										
Qy	1431	tggcttcagaggtgggttag--aagattctaatctcagggctctatgataaaaaaagaac 1488										
Db	720	CACCTTCNAAGGTGGAAGGAAATTTCTGCTTTTCAGGCACCACTGTTTGAATAAAAAA 779										
Qy	1489	ggtcagatattctcaaaa 1506										
Db	780	AACCAACCCGTTNCTTAAA 797										

Db	361	TACATTGGCTCTTTACCTCTTCCACATTGCTGGAGCTTACCTTTTGAACCTTGAATCATCTA	420
Qy	1131	ggactctcttttgggtactgcattatttggtaaatcatttcccacatgtgcggcctg	1190
Db	421	GGACTTGCTCTCTGGTGCACATATTATTTGTTGCAATTTCTTTTCCACATTTCCCGCCTG	480
Qy	1191	ttttactttagtgtaaaaagtcacgaagaagcatctctgtggccattgtgtttatc	1250
Db	481	TTTTTATTTTAGCAATGAAGAAAGTATCAGAAGAGATTTCCTGTGGCGAGTCTCTTTTGTGT	540
Qy	1251	ttgggtagactgtgaccttcaattgtttccggtactcactgttgggtttcaoctggctgga	1310
Db	541	TTGGGAANATTCTGACTTTAAATCTTTCAATACTGACTGTGTGTTTTGGCTTGGCNA	600
Qy	1311	tcgcagaatcggaatcctgctgccttacttggaaatgtaaatgtgttggcagctaaatt	1370
Db	601	ACANAANAATCAGAAACTGGATTTCAGTACTTGGAAACTTCATGTGTTT-ACTGTGTTAAATC	659
Qy	1371	gctgttctgtctgtccadgtgcgaogatccaagctcacgtaacatgggaacttaattactctc	1430
Db	660	GCTGTTCTGGGATCCATTTCGTTTNCAGGCATTTATAATGTGGAAATTCATTAATTTT	719
Qy	1431	tggcttcagaggtgggtag--aagattctaatatctcggctctatgtatgaaaagaaac	1488
Db	720	CACCTTCNAAGGTGAAGGAAAAATTTCTGCTTTTCAGGCACCAACTTGTGTGAAAAAAA	779
Qy	1489	ggtcgagatctcttaaaa	1506
Db	780	AACCACCCGTTNCTTAAAA	797

RESULT 15

BM459490
LOCUS
DEFINITION AGENCOURT_641971 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5534104
5', mRNA sequence.
ACCESSION
VERSION BM459490.1 GI:18508530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: LLM12220 row: a column: 17
High quality sequence stop: 696.

FEATURES

source	Location/Qualifiers
1..910	
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:5534104"
	/clone_lib="NIH_MGC_71"
	/tissue_type="leiomyosarcoma"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."
BASE COUNT	226 a. 222 c 207 g 255 t
ORIGIN	

Query Match 21.6%; Score 495.2; DB 10; Length 910;
Best Local Similarity 82.2%; Pred. No. 6.3e-91;
Matches 618; Conservative 0; Mismatches 128; Indels 6; Gaps 4;

QY	465	ggaagcaccacatgggctcgttaagaagacagcaccagacccccccggtctcagccag	524
Db	128	GGCGCCCCACATGGCGATTCCGAAGAAAGACCAAGAGCGCCCCCAGTGTGAGCCAC	187
QY	525	gaattcatcctgcagaatcatcgcgacatcgtctcctgctggggatgtctctcctgctg	584
Db	188	GAATTCGTCTGCAGATCAGCGCGACATCGTCCTGTGTGGGATGGTCTTCTCTGTG	247
QY	585	gggtctgtgttcgagggaacagcagagatccatcctgtttctcactcttcagcacagt	644
Db	248	GGGCTCATGTTTGAGATAACGGCAAAAGCTCTATCATTTTGTGTACTCTTCAGTACAAT	307
QY	645	gttgcgtccctgcagcagaggaaacaagccacggggtccaaagtccctctattatggt	704
Db	308	GTCACCTCCAGCAACAGAGAAAGACTACTGAATCAGTGTCCCTTTATTACTATGGC	367
QY	705	gtcaagatttggccaggggtttctctacatcgtggtggcaatcattatcctgccaca	764
Db	368	ATCAAGATTTGGTACTGTTTCTTCTACATGCTAGTGGCGATAATTATTCATGCCGTA	427
QY	765	attcagaatatgtgtggataaaattacaagagaatgcagttcacccaagcgaaacaa	824
Db	428	ATTCAAGAGTATATTTGGATAAAATTAACAGGCGAATGCACCTCTCCAAAACAAACAC	487
QY	825	aacaagtttaacagagctggtcagtttagtgtgtctactttttttttttttttttttgggc	884
Db	488	AGCAAGTTTAATGAATGGGTGACGTAGTGCCTTACCTTTTTCCTGTGTGTTGGGC	547

QY	885	acattcattttaatctctgaaactgctgtcagaccccaactctcttatatggaaggtcgt	944
Db	548	ACATTCAATTCATCTCTGAAACTACATCTCAGACCCCACTATCTTATGGAGGCTTAT	607
QY	945	ccccatagcatgatgacatttcaaatgaagtgtttctacatatccccagttggttactg	1004
Db	608	CCCATAAACCTGATGACATTTCAAATGAAGTTTTTCTACATATCACAGCTGGCTTACTGG	667
QY	1005	tttcagtcttttctgaactctacttccagaaaaacacacacacacacacacacacacac	1064
Db	668	CTTCATGCTTTTCTGAACTCTACTTCCAGAAAAACAAAAAGAGATATCTCTCGTCAG	727
QY	1065	cttgtctacatattggtctctcaccctcttcacatt-actggagcttattctctt-gtacttga	1122
Db	728	CTTGTCTACATTGGCCTTTACCTCTTTTCACATTGGCTGGAGCCTACCTTTTGGAACTTGA	787
QY	1123	atcat-ttgggaactctctcttgggtactgcattattttgtgaattac---tttccac	1178
Db	788	ATCATCTAGGAGCTTGTCTCTGGCGCTACATTTATTTGGTTGATTTCTCTTTCCCAT	847
QY	1179	atgtgggctgttttacttttagtgatgaaaa	1210
Db	848	TTTCCCGCCTGTGTTTATTATTTAGCAATGAAAA	879

Search completed: September 6, 2002, 13:08:10
Job time: 46629 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:09:47 ; Search time 74.4 Seconds
(without alignments)
550.890 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKRKSTKPPVLSQEFIL.....VGVTNSNRVDCPPKRRKRS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1920	100.0	369	21	AAV98147 Human WAR-1 amino
2	1920	100.0	369	22	ABG12234 Novel human diagno
3	1920	100.0	369	22	AA870895 Human WAR-1 protei
4	1919	99.9	369	22	AAW93265 Human polypeptide,
5	1463	76.2	363	21	AAV98146 Rat WAR-1 amino ac
6	1463	76.2	363	22	AA870696 Rat WAR-1 protein
7	1392.5	72.5	416	21	AA843601 Human cancer assoc
8	1129.5	58.8	304	20	AAV48434 Human prostate can
9	815	42.4	384	22	AAW95708 Human reproductive
10	602.5	31.4	368	22	ABB65847 Drosophila melanog
11	602.5	31.4	368	22	ABB66114 Drosophila melanog

12	602.5	31.4	1575	22	ABB67362 Drosophila melanog
13	493	25.7	125	21	AAG00189 Human secreted pro
14	144	7.5	411	17	AA86810 Saccharomyces cere
15	143	7.4	394	20	AAV00876 Human LAPH-1 prote
16	143	7.4	394	22	AAW78909 Human protein sequ
17	143	7.4	394	22	AAW93884 Human protein sequ
18	138	7.2	191	21	AA824489 Arabidopsis thalia
19	127	6.6	310	21	AAG41779 Arabidopsis thalia
20	117.5	6.1	488	22	AAW42028 Human polypeptide
21	117.5	6.1	536	22	AAW79893 Human protein sequ
22	116.5	6.1	280	21	AAW44197 Arabidopsis thalia
23	116.5	6.1	296	21	AAW44196 Arabidopsis thalia
24	116	6.0	380	20	AAV00877 Human LAPH-2 prote
25	116	6.0	380	22	AAW88560 Human hydrophobic
26	115	6.0	254	21	AAG41780 Arabidopsis thalia
27	114.5	6.0	181	22	AAU17201 Novel signal trans
28	114	5.9	266	21	AAW43806 Human cancer assoc
29	113	5.9	158	22	AAU17576 Novel signal trans
30	106.5	5.5	332	22	AAW72076 Human olfactory re
31	105.5	5.5	484	22	AAU34864 E. coli cellular p
32	105.5	5.5	816	20	AAV27339 Group B Streptococ
33	104.5	5.4	150	21	AAW34938 Human secreted pro
34	102	5.3	311	22	AAU00782 Human apoptosis pr
35	102	5.3	692	22	AAW35236 Human glycine tran
36	101	5.3	638	22	AAW47065 Glycine transporte
37	101	5.3	692	22	AAW35235 Human neurotransmi
38	100.5	5.2	339	22	ABB60006 Drosophila melanog
39	100	5.2	414	22	AAW23517 Human EST encoded
40	100	5.2	642	21	AAW81943 Human GLYTLIKE pro
41	100	5.2	642	21	AAW81943 Human GLYTLIKE pro
42	100	5.2	1898	22	ABG25514 Novel human diagno
43	99.5	5.2	313	21	AAG33404 Zea mays protein f
44	98.5	5.1	577	22	ABG29227 Novel human diagno
45	97.5	5.1	262	22	ABB66068 Drosophila melanog

ALIGNMENTS

RESULT 1
AAV98147
ID AAV98147 standard; Protein; 369 AA.
XX
AC AAV98147;
XX
DT 22-AUG-2000 (first entry)
XX
DE Human WAR-1 amino acid sequence.
XX
KW Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; human.
XX
OS Homo sapiens.
XX
FN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
XX (SUMU) SUMITOMO PHARM CO LTD.
XX
XX Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
XX Okuyama H;
XX WPI; 2000-317980/27.
XX N-PSDB; AAA38013.
XX
XX Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
XX proliferation for use in treatment and diagnosis of cancer including
XX sarcomas of high malignancy -

XX Claim 1; Fig 2; 89pp; Japanese.
XX This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridise to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
XX Sequence 369 AA;

Query Match 100.0%; Score 1920; DB 21; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRSTKSTNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
Db 1 mglrskstknppvlsqefilqnhadivscvgmffllglvfegetaeasivfltlqhsavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
Db 61 aaeqatgsksllyygvkdlatvffymvliiiahiatqeyvldkinkrmqftkakqnf 120
QY 121 ESGQSFVYFFSCIWGTFLISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWFAHAF 180
Db 121 esgqsfvfyffsciwgtflisencldptliwkarphsmftfomkffiyisolaywfhaf 180
QY 181 PELYFQKTKKODIPQOLVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
Db 181 pelyfqtktkodipqolvylghlflhitgaylllynhgllylhlyfvelshmcglfy 240
QY 241 FSDEKYQKGISLWAIIVFLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKTAV 300
Db 241 fsdekyqgislwaiivflgrvltivsvltvgtfhlagsqnrnpdaltgnnvlaakia 300
QY 301 LSSCTIQAYVTWNLITLWLRQWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
Db 301 lssctiqayvtwnlitlwlqrwvedsnioascmkkrrsrrkrtrtengvgvetsnrvc 360
QY 361 PPKRKEKSS 369
Db 361 ppkrkekss 369

RESULT 2
ABG12234
ID ABG12234 standard; Protein; 369 AA.
XX AC ABG12234;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #12225.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
DR N-PSDB; AAS76421.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID No 42593; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 369 AA;

Query Match 100.0%; Score 1920; DB 22; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRSTKSTNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
Db 1 mglrskstknppvlsqefilqnhadivscvgmffllglvfegetaeasivfltlqhsavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKQNFN 120
Db 61 aaeqatgsksllyygvkdlatvffymvliiiahiatqeyvldkinkrmqftkakqnf 120
QY 121 ESGQSFVYFFSCIWGTFLISENCLSDPTLIWKARPHSMFTFQMKFFYISOLAYWFAHAF 180
Db 121 esgqsfvfyffsciwgtflisencldptliwkarphsmftfomkffiyisolaywfhaf 180
QY 181 PELYFQKTKKODIPQOLVYIGLHFLHITGAYLLYLNHGLLLVHLYFVELLSHMCGLFY 240
Db 181 pelyfqtktkodipqolvylghlflhitgaylllynhgllylhlyfvelshmcglfy 240
QY 241 FSDEKYQKGISLWAIIVFLGRVLTIVSVLTVGFHLAGSQNRNPDALTGNNVLAAKTAV 300
Db 241 fsdekyqgislwaiivflgrvltivsvltvgtfhlagsqnrnpdaltgnnvlaakia 300
QY 301 LSSCTIQAYVTWNLITLWLRQWVEDSNIOASCMKKRSRKRRTENGVGVTENRVD 360
Db 301 lssctiqayvtwnlitlwlqrwvedsnioascmkkrrsrrkrtrtengvgvetsnrvc 360
QY 361 PPKRKEKSS 369
Db 361 ppkrkekss 369
RESULT 3
AAB70695
ID AAB70695 standard; Protein; 369 AA.

XX AAB70695;
XX 17-MAY-2001 (first entry)
XX Human WAR-1 protein sequence SEQ ID NO:1.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; anti-allergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
XX Homo sapiens.
XX WO200114582-A1.
XX 01-MAR-2001.
XX 17-AUG-2000; 2000WO-JP05488.
XX 20-AUG-1999; 99JP-0234764.
XX (SUMU) SUMITOMO PHARM CO LTD.
XX Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX WPI: 2001-202940/20.
XX N-PSDB; AAF74780, AAF74782.
XX Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell
XX
XX Claim 2; Page 58-60; 79pp; Japanese.
XX
CC The present invention describes a screening method for secretory and
CC membrane proteins consisting of transformation of a cell with separate
CC expression vectors for the sense and antisense RNA of DNA encoding an
CC endoplasmic reticulum membrane protein participating in endoplasmic
CC reticulum transport of proteins. Also described are: (1) secretory and
CC cell membrane proteins identified by the screening method; (2) drug
CC compositions containing these proteins; (3) host cells transformed by
CC the separate expression vectors of the method; and (4) the preparation
CC of secretory and cell membrane proteins by culture of the transformants.
CC The method can be used for the identification and preparation of
CC proteins for use in the treatment and prevention of diseases such as
CC cancer, disorders of the nervous system, immune disorders (including
CC allergies and rheumatism) and skeletal disorders. The present sequence
CC represents a specifically claimed human WAR-1 protein from the present
XX invention.
XX
SQ Sequence 369 AA;

Query Match 100.0%; Score 1920; DB 22; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.7e-214;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGTAESAIVFLTLQHSVAVP 60
DB 1 mglrkkstknppvlsqefilqnhadivscvgmffllglvfgetaesivfltlqhsavp 60

QY 61 AAEQATGSKSLYYGVKDLATVFFMLVAIIHATIQEYVLVDKINKRMQFTAKQNKFN 120
DB 61 aaeqatgskslyygvkdlatvffmlvaaiihati qeyvlvd kinkrmqftakqknfn 120

QY 121 ESGQFVFFFCITGTFILISNCLSDPTLIWKARPHSNMTFQMKFFYISQLAYWFAH 180
DB 121 esqgfsvfffcitgtf ilisnclsdptliwkarphsnmtf qmkffysqlaywfhaf 180

QY 181 PELYFKTKQDIPROLVYITGLHFLHITGAYLLYLNHLGLLLEVLVHYFVELLSHMGGLFY 240
DB 181 pelyfktkqdiprolvyitglhflhitgayll ylnhlgl llllevlvhyfvel lshmgglfy 240

DB 181 pelyfktkkdiprqlvyiglhlhifhitgayll ylnhlgl llllvlyhfvellshmgglfy 240
QY 241 FSDEKYOKGISLWAIIVFLIGRLVTLIVSVLTGVGPHLAGSQNRNPDALTGNVNLAAKIAV 300
DB 241 fdekyokgislwaivfligrlvtli vsvltgv gphlagsqnrnpdaltgnvnvnl aakiav 300
QY 301 LSSSCTTQAYVTWNLTITLWLRQWVEDSNIOASCMKKRKRSSKKRRTENGVGVEYSNRVDC 360
DB 301 lsssectiqayvtwnltitlwlrqwvedsn i oascmk krrsskkrrtengvgvetsnr vdc 360
QY 361 PPKRKEKSS 369
DB 361 ppkrkekss 369

RESULT 4
AAM93265
ID AAM93265 standard; Protein; 369 AA.
XX
AC AAM93265;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2725.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
XX WPI: 2001-524255/58.
XX N-PSDB; AAK94181.
XX
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation
PT
PS Claim 8; SEQ ID NO 2725; 1380pp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 369 AA;

Query Match 99.9%; Score 1919; DB 22; Length 369;
Best Local Similarity 99.7%; Pred. No. 3.5e-214;
Matches 368; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFEGTAESAIVFLTLQHSVAVP 60

```
Db 1 mglrkstknpvlsqefilqhadivscgmffllglvfegtaeasivfltlqhsavp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 aaeqatgskslYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
QY 121 ESGQSFVFFSCIGTFTLISENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYWFHAF 180
Db 121 esgqsfvffscigTFTLISENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYWFHAF 180
QY 181 PELYQKTKQDIPROLVYIGLHLPHITGAYLLYLNLHGLLLLVLFYVELLSHMGGLFY 240
Db 181 pelyqktkkdiprqlvYIGLHLPHITGAYLLYLNLHGLLLLVLFYVELLSHMGGLFY 240
QY 241 FDEKQKGISLWAIIVFILGRVTLIVSVLTGCFHLAGSONRNPDALTCGNVNVLAAKIAV 300
Db 241 fdekqkgislwaiVfIlgrVtlivsvltGcfhlagsnRnpdaltcgnvnlvLaakiav 300
QY 301 LSSSCTIQAYVTWNLTITLQWRVEDSNIAQSCMKKKRSRKKRTENGVGVSNSRVDC 360
Db 301 lssscTiQayvtwnltitlQwrvedsniaqscmkkkrsrskkrtengvgvsnsrvdc 360
QY 361 PPKRKEKSS 369
Db 361 pprkrekss 369

RESULT 5
AAY98146
ID AAY98146 standard; Protein; 363 AA.
AC AAY98146;
XX
DT 22-AUG-2000 (first entry)
XX
DE Rat WAR-1 amino acid sequence.
KW Endoplasmic reticulum; WAR-1; cancer cell proliferation inhibitor;
KW diagnosis; cancer; sarcoma; rat.
XX
OS Rattus norvegicus.
XX
PN WO200022123-A1.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-JP05631.
XX
PR 13-OCT-1998; 98JP-0290711.
XX
PA (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Yoshima T, Komiya K, Tojo S, Nemoto K, Ishikawa H;
PI Okuyama H;
XX
DR WPI; 2000-317980/27.
DR N-PSDB; AAA38012.
XX
PT Endoplasmic reticulum protein WAR-1 which inhibits cancer cell
PT proliferation for use in treatment and diagnosis of cancer including
PT sarcomas of high malignancy -
XX
PS Claim 1; Fig 2; 89pp; Japanese.
XX
CC This sequence represents an endoplasmic reticulum protein (WAR-1) amino
CC acid sequence. The invention includes rat and human WAR-1 sequences,
CC expression vectors containing the DNA, cells transformed with the
CC expression vector, antibodies against WAR-1, and probes and primers which
CC hybridize to the DNA encoding WAR-1. The WAR-1 protein inhibits the
CC proliferation of cancer cells, and is used in the treatment and diagnosis
CC of cancers including highly malignant sarcomas.
XX
```

```
SQ Sequence 363 AA:
Query Match 76.2%; Score 1463; DB 21; Length 363;
Best Local Similarity 75.8%; Pred. No. 3.7e-161;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;

QY 1 MGLRKKSTKNPPVLSQEFILQHADIIVSCVGMFFLLGLVFEFTAASIVFLTLQHSVAVP 60
Db 1 mglrkknarppvlshefvmqhadmvsqgmffvlglmfegtaeasivfltlqhgvvvp 60
QY 61 AAEQATGSKSLYYGVKDLATVFFVMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 -aeglpsgsrtilyhygvkdlatvffymlvaiiiahatiqeyvldklsrrlqltkgkqnkn 119
QY 121 ESGQSFVFFSCIGTFTLISENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYWFHAF 180
Db 121 esgqlsvfYVIGVGLWGMILLASCLSDPTLIWksphmmmtfQmkffYisqlayfhshf 179
QY 181 PELYQKTKQDIPROLVYIGLHLPHITGAYLLYLNLHGLLLLVLFYVELLSHMGGLFY 240
Db 180 pelyfqkvkqdiPgqliyIghlfigayllylnhlgllllmlhyavellssvcsllly 239
QY 241 FDEKQKGISLWAIIVFILGRVTLIVSVLTGCFHLAGSONRNPDALTCGNVNVLAAKIAV 300
Db 240 fgderYqkglswpivfIsgrlvtlivsvvtvgqlhlagt-nrnngnalsgnvnvLaakiav 298
QY 301 LSSSCTIQAYVTWNLTITLQWRVEDSNIAQSCMKKKRSRKKRTENGVGVSNSRVDC 360
Db 299 lssscsiqYitwtltvtwqlrWledanlhv-cgrkrirsr-srkgtgn--gvnparids 354
QY 361 PPKRKEKS 368
Db 355 pppkkkka 362

RESULT 6
AAB70696
ID AAB70696 standard; Protein; 363 AA.
AC AAB70696;
XX
DT 17-MAY-2001 (first entry)
XX
DE Rat WAR-1 protein sequence SEQ ID NO:2.
XX
KW WAR-1; protein screening; endoplasmic reticulum membrane protein;
KW endoplasmic reticulum membrane transportation; secretory protein;
KW cell membrane protein; cytosolic; CNS active; antiallergic; cancer;
KW antirheumatic; nervous system disorder; immune disorder; allergy;
KW rheumatism; skeletal disorder.
XX
OS Rattus sp.
XX
PN WO200114582-A1.
XX
PD 01-MAR-2001.
XX
PF 17-AUG-2000; 2000WO-JP05488.
XX
PR 20-AUG-1999; 99JP-0234764.
XX
PA (SUMU ) SUMITOMO PHARM CO LTD.
XX
PI Tohdoh N, Okuyama H, Imamura M, Ishikawa H, Nemoto K;
XX
DR WPI; 2001-202940/20.
DR N-PSDB; AAF74781.
XX
PT Transformation of a cell with separate vectors expressing the sense and
PT antisense strands of WAR-1 DNA for screening secretory and membrane
PT proteins expressed by the cell -
XX
```


PS Claim 3; Page 60-62; 79pp; Japanese.

XX The present invention describes a screening method for secretory and membrane proteins consisting of transformation of a cell with separate expression vectors for the sense and antisense RNA of DNA encoding an endoplasmic reticulum membrane protein participating in endoplasmic reticulum transport of proteins. Also described are: (1) secretory and cell membrane proteins identified by the screening method; (2) drug compositions containing these proteins; (3) host cells transformed by the separate expression vectors of the method; and (4) the preparation of secretory and cell membrane proteins by culture of the transformants.

CC The method can be used for the identification and preparation of proteins for use in the treatment and prevention of diseases such as cancer, disorders of the nervous system, immune disorders (including allergies and rheumatism) and skeletal disorders. The present sequence represents a specifically claimed rat WAR-1 protein from the present invention.

XX Sequence 363 AA;

SQ

Query Match 76.2%; Score 1463; DB 22; Length 363;
Best Local Similarity 75.8%; Pred. No. 3.7e-161;
Matches 279; Conservative 43; Mismatches 40; Indels 6; Gaps 5;

QY 1 MGLRRKTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFECTAEASIVFLTLQHSVAVP 60
DB 1 mglrrktnppvlshefmvqhdamvcvgmffvlgmfegtaemsivfltlqhgvvvp 60

QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONKFN 120
DB 61 -aeglpsgsrtlyhygvkdlatvffymvliiahiatiqeyvldklsrrlqlckgkqnln 119

QY 121 ESGQFSVYFFSCINGTFTLLISENCLSDPTLIWKARPHSMFTFQMKFFYSOLAYWFHAF 180
DB 120 eagqlsviyvsglwgmlsasencldptllwksqphnmftgmkfifysqlaywfhsf 179

QY 181 PELYFQKTKKODIPQLVYIGLHLEPHITGAYLLYNHLGLLLVLYFVELLSHMCGLFY 240
DB 180 pelyfqkvrkqdpqglylgllhfhigayllynhlglllmhyavellssvcsily 239

QY 241 PSDEYKQGISLWALVFLIGRLVTLIVSVLTVGFHLAGSQNRNPALTCGNVNLAAKTAV 300
DB 240 fdeyqkglslwplvifsgrlvtlsvvtvghlagt-nrngnalsgnvnvlaakiav 298

QY 301 LSSSCCTIOAYVTWNLITLWLRWVEDSNIQASCMKKRSRSEKKTENGVGVTENRVDC 360
DB 299 lssscslqvvtwtltvqlrledanlhv-cgrkrrsr-srkgten--gvnparids 354

QY 361 PPKRREKS 368
DB 355 ppkkteka 362

RESULT 7
ID AAB43601 standard; Protein; 416 AA.
XX AAB43601;
AC AAB43601;
XX 08-FEB-2001 (first entry)
XX Human cancer associated protein sequence SEQ ID NO:1046.
XX Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;

KW neurological disease; drug screening.

XX Homo sapiens.

OS WO200055350-A1.

XX 21-SEP-2000.

PD 08-MAR-2000; 2000WO-US05882.

PF 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

PI WPI; 2000-587533/55.

XX N-PSDB; AAC77810.

DR Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer

PT Claim 11; Page 1634-1636; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given in AAB43398 to AAB44239. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulnerable; immunomodulator; antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiinflammatory; antithyroid; antiallergic; antibacterial; coagulant; dermatological; neuroprotective; cardiac; thrombolytic; vasotrophic; vasotrophic; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present invention.

XX Sequence 416 AA;

SQ

Query Match 72.5%; Score 1392.5; DB 21; Length 416;
Best Local Similarity 71.4%; Pred. No. 7e-153;
Matches 267; Conservative 44; Mismatches 58; Indels 5; Gaps 2;

QY 1 MGLRRKTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFECTAEASIVFLTLQHSVAVP 60
DB 43 mairkkstksppvlshefvqlqhadvscvnmfvllgmfeitakasiifvltqynvtlp 102

QY 61 AAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONKFN 120
DB 103 ateeqatesvlyyygiklatvffymvliiahiatiqeyvldklnrrmhfskthskfn 162

QY 121 ESGQFSVYFFSCINGTFTLLISENCLSDPTLIWKARPHSMFTFQMKFFYSOLAYWFHAF 180
DB 163 esgqlsalfylfacwgtfllisenysdptliwrayphnltmfqmkffysqlaywlfhaf 222

QY 181 PELYFQKTKKODIPQLVYIGLHLEPHITGAYLLYNHLGLLLVLYFVELLSHMCGLFY 240
DB 223 pelyfqtkkdedipqlylgllhfhigayllynhlgllvlvhyfvefifhsrlyf 282

QY 241 PSDEYKQGISLWALVFLIGRLVTLIVSVLTVGFHLAGSQNRNPALTCGNVNLAAKTAV 300
DB 283 fsnekyqkgfslwavlflvrltllsvltvfgfqlaraenqkldfstgnfnlavriav 342

DR WPI; 2000-500381/45.
XX N-PSDB; AAC00195.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 4270; 71pp + CD-ROM; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
XX Sequence 125 AA;

Query Match 25.7%; Score 493; DB 21; Length 125;
Best Local Similarity 76.6%; Pred. No. 5e-49;
Matches 95; Conservative 15; Mismatches 14; Indels 0; Gaps 0;
QY 1 MGLRKKSTKNPVLVSQEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
DB 1 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 60
1 mairkkstkppvlshfvlqnhadivscvamvflglmfeltakasiifvtlqynvtip 60
QY 61 AAEQATGSKSLYYGVKDLATVFFVFMVAIIHATIQEYVLDKINKRMQFTKAKQNKFN 120
DB 1 :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 120
61 ateeqatesvslyygiqlatvffvfmvliiiaihavieymidkinnrmxfsktxhskn 120
QY 121 ESQ 124
DB 121 esgq 124
|||||

RESULT 14
AAR86810
ID AAR86810 standard; Protein; 411 AA.
XX
XX AAR86810;
AC
DT 28-MAY-1997 (first entry)
XX
XX Saccharomyces cerevisiae LAG1 protein.
XX
XX Saccharomyces cerevisiae; LAG1; life-span limiting domain;
KW life-span extending domain; stress tolerance; longevity;
KW recombinant protein production.
XX
XX Saccharomyces cerevisiae.

Key Location/Qualifiers
FT Domain 1..190
FT /note= "life-span limiting domain"
FT 194..411
FT /note= "life-span extending domain"
XX
XX WO9533834-A1.
XX
XX 14-DEC-1995.
XX
XX 02-JUN-1995; 95WO-US06725.
XX
XX 08-NOV-1994; 94US-0336031.
PR 03-JUN-1994; 94US-0253875.

XX (RESE) RESEARCH CORP TECHNOLOGIES INC.
XX Jazwinski SM;
XX
XX WPI; 1996-040238/04.
DR N-PSDB; AAT07263.
XX
XX Eukaryotic LAG1 gene and protein - controls longevity, stress
PT tolerance and reproductive capacity of eukaryotic cells, for
PT improved prodn. of recombinant proteins.
XX
XX Claim 23; Page 100-102; 154pp; English.
XX
XX The Saccharomyces cerevisiae LAG1 gene product is composed of a
CC life-span limiting domain (see AAR86812) and a life-span extending
CC domain (see AAR86811). Overexpression of the LAG1 gene in older
CC cells has a rejuvenating effect, which not only increases
CC cellular life span, but also reproductive capacity and cellular
CC tolerance to stress factors such as starvation and low pH. These
CC cells may be used for the production of recombinant proteins. By
CC increasing the life span of recombinant cells, the need for
CC overexpression of recombinant gene products is avoided, and
CC therefore any subsequent adverse effects on the host cell.
XX
XX Sequence 411 AA;

Query Match 7.5%; Score 144; DB 17; Length 411;
Best Local Similarity 22.6%; Pred. No. 1.1e-07;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;
QY 33 FELLGLVFEGETAEASIVFLTLQHSVAVP-----AAPEQATGSKSLYYGVKDLATVFFYM 87
DB 88 ffilvcvy-----sayflsgrntesnlhmfvaisyqvdgtds-yakgikdlisffvfm 140
QY 88 LVAIIHATIQEYVLDKINKRMQFTKAKQNKFNESQFSVFYF-FSCINGTFTILI-SENC 145
DB 141 ifftflreflmdvirpftvynvtsehrqkrmleqmaifycvgvgpgfglymyhdsdlw 200
QY 146 LSDPTLIWKARPHSMMTFQMKFFYISQLAYWFHAFPELYFOKTKODIPROLYVIGL-HL 204
DB 201 lftkpmrytpvntpnflkifvgaaqwaqgacvvlqlkprkdykelvfhivtl 260
QY 205 FHTGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLFY----- 240
DB 261 lliwssyvfhtkmglaitytmdvdsfflslsktlnylnsvftpfvfglffvfiylrhv 320
QY 241 -----FSDEKYOKGIGSLWAIVFILGRVLTILI 266
DB 321 vnirilwsvltfrehgnyvlnfatgqkwcwisl-pivfvliaalqlv 367

RESULT 15
AAY00876
ID AAY00876 standard; Protein; 394 AA.
XX
XX AAY00876;
AC
DT 21-MAY-1999 (first entry)
XX
XX Human LAPH-1 protein sequence.
DE
DE Longevity-assurance protein homologue; LAPH-1; human; cell proliferation;
KW longevity-assurance protein homologue; LAPH-2; signal transduction;
KW cell cycle regulation; apoptosis; cellular homeostatic pathway; aging;
KW cancer; inflammation; autoimmune disease; infection;
KW neurodegenerative disorder.
XX
XX Homo sapiens.
XX
XX WO9906558-A1.
XX

Search completed: September 6, 2002, 17:09:48
Job time: 8277 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:10:38 ; Search time 28.82 Seconds
(without alignments)
312.736 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKKSTKNPPVLSQEFIL.....VGVERSNRVDPPPKRKES 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCRU5_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	144	7.5	411	2	US-08-336-031-2
2	144	7.5	411	2	US-08-902-853-7
3	144	7.5	411	5	PCT-US95-06725-2
4	143	7.4	394	2	US-08-902-853-1
5	128	6.7	387	2	US-08-902-853-6
6	116	6.0	380	2	US-08-902-853-3
7	102	5.3	687	3	US-08-834-467-2
8	102	5.3	687	4	US-09-396-177-2
9	94.5	4.9	696	4	US-07-757-342D-4
10	93	4.8	638	1	US-08-295-814E-13
11	93	4.8	638	1	US-08-240-783B-2
12	93	4.8	638	4	US-09-084-813-2
13	93	4.8	638	4	US-09-343-361-13
14	93	4.8	638	5	PCT-US92-09662-2
15	86	4.5	193	2	US-08-336-031-6
16	86	4.5	193	5	PCT-US95-06725-6
17	86	4.5	2020	1	US-07-551-531-2
18	86	4.5	2485	5	PCT-US94-00198-1
19	86	4.5	2485	5	PCT-US94-00198-2
20	86	4.5	2818	1	US-08-510-284-1
21	86	4.5	2818	1	US-08-411-389-2
22	86	4.5	2818	2	US-08-449-933-2
23	86	4.5	2818	4	US-07-966-049A-2
24	86	4.5	2818	4	US-09-542-331-2
25	84	4.4	1956	4	US-08-843-417-2
26	83.5	4.3	362	2	US-08-902-853-5
27	83	4.3	317	2	US-08-790-572-3

28	83	4.3	317	2	US-09-213-398-3	Sequence 3, Appli
29	83	4.3	764	4	US-07-741-453A-60	Sequence 60, Appl
30	81	4.2	764	4	US-09-235-451-36	Sequence 36, Appl
31	80.5	4.2	363	3	US-09-046-086-2	Sequence 2, Appli
32	80	4.2	429	2	US-08-677-049-5	Sequence 5, Appli
33	80	4.2	1956	4	US-08-843-417-10	Sequence 10, Appl
34	78.5	4.1	317	2	US-08-619-362A-8	Sequence 8, Appli
35	77.5	4.0	494	1	US-08-689-974-3	Sequence 3, Appli
36	77.5	4.0	494	3	US-09-058-376-3	Sequence 5, Appli
37	77.5	4.0	1036	2	US-08-720-484A-5	Sequence 5, Appli
38	77.5	4.0	1036	4	US-08-953-823A-5	Sequence 5, Appli
39	77	4.0	395	3	US-08-781-891-73	Sequence 73, Appl
40	77	4.0	1269	3	US-08-781-891-74	Sequence 74, Appl
41	77	4.0	1432	3	US-08-781-891-71	Sequence 71, Appl
42	76.5	4.0	637	3	US-08-072-064-8	Sequence 8, Appli
43	76	4.0	617	1	US-07-879-617A-11	Sequence 11, Appl
44	76	4.0	617	1	US-08-301-722A-3	Sequence 3, Appli
45	76	4.0	617	1	US-08-240-783B-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-336-031-2
; Sequence 2, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-336-031-2

Query Match 7.5%; Score 144; DB 2; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;
Qy 33 FFLGLVFEPTAEASIVFTLQHSVAVP-----AAEQATGSKSLYYGVKDLATVFFYM 87

Db 88 FFILVVCY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140
QY 88 LVAIIITHATIQEYVLDKINKRMQFTRAKQKNFESGQFSVFYF-FSCIMGTFFILI-SENC 145
Db 141 IFFTLREFLMDVIRPFTVYLVNTSEHROKRMLEQMYAIFYCGVSGPFGGLYIMYHSDLW 200
QY 146 LSDPTLLWKARPHSMFTOMKFFFYISOLAYWFAHFAFPPELYFQKTKKODIPROLVYIGL-HL 204
Db 201 LFKTKPMRYTPVITNPFLEKFIYLGQAFAWAOACVVLVQLQEKPRKDYKELVFHHIVTL 260
QY 205 FHITGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLEY----- 240
Db 261 LLWSSVVFHTKMGLAITYTMDVDFSLSKTLNLYNSVFTPFVGLVFFVFIYLRHV 320
QY 241 -----FSDEKYQKGISLWAIWVFIILGRVTLI 266
Db 321 VNIRILSVLTERHGNVYLNFEATQYKWCISL-PIVFLVIAALQLV 367

RESULT 2

US-08-902-853-7
; Sequence 7, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 541568
; US-08-902-853-7

Query Match 7.5%; Score 144; DB 2; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFILGLVFECTAASIVFLTLQHSVAVP-----AAEQATGSKSLYYGYVKDLATVFFYM 87
Db 88 FFILVVCY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140
QY 88 LVAIIITHATIQEYVLDKINKRMQFTRAKQKNFESGQFSVFYF-FSCIMGTFFILI-SENC 145
Db 141 IFFTLREFLMDVIRPFTVYLVNTSEHROKRMLEQMYAIFYCGVSGPFGGLYIMYHSDLW 200
QY 146 LSDPTLLWKARPHSMFTOMKFFFYISOLAYWFAHFAFPPELYFQKTKKODIPROLVYIGL-HL 204
Db 201 LFKTKPMRYTPVITNPFLEKFIYLGQAFAWAOACVVLVQLQEKPRKDYKELVFHHIVTL 260
QY 205 FHITGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLEY----- 240
Db 261 LLWSSVVFHTKMGLAITYTMDVDFSLSKTLNLYNSVFTPFVGLVFFVFIYLRHV 320
QY 241 -----FSDEKYQKGISLWAIWVFIILGRVTLI 266
Db 321 VNIRILSVLTERHGNVYLNFEATQYKWCISL-PIVFLVIAALQLV 367

RESULT 3

PCT-US95-06725-2
; Sequence 2, Application PC/TUS9506725
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06725
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875 & 08/336,031
; FILING DATE: 03-JUN-1994 & 08-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06725-2

Query Match 7.5%; Score 144; DB 5; Length 411;
Best Local Similarity 22.6%; Pred. No. 3.7e-08;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFILGLVFECTAASIVFLTLQHSVAVP-----AAEQATGSKSLYYGYVKDLATVFFYM 87
Db 88 FFILVVCY-----SAYFLSGNRTESNPLHMFVAISVQVDGTD-SAYGKIKDLSFFFFYM 140

```

QY 88 LVAIIHTATIQEVLDKINKRMQFTKAKONKFNESQGSVFYF-FSCIWGTFLILI-SENC 14
DB 141 IFPTFLRELMQVIVRPFTVYLNVITSEHRQKRMLEQMYAIFYCGVSGPGLYIMYHSDLW 200
QY 146 LSDPTLIWAKRPHSMFTQMKFEYISOLAYWEHAPPELKYOKTKKQDIPROLVYIGL-HL 204
DB 201 LFTKPMIRYPIVITNPDLFKIFYLQGAFAAQACVLVLQLEKPKKDYKELVFHHIIVTL 260
QY 205 FHTGAYLLYLNLHGLLLLV-----LHYFVELLS-HMCGILFY----- 240
DB 261 LLWSSYVVFHFTKGLAIYITWDVDFFLSKLTNLYLSNVTPFVFGLEVFVFIYLRHV 320
QY 241 -----FSDBKYOKGISLNAIVEFILGRVLTL 266
DB 321 VNIRILWSVLTEFRHGNVVLNFATQQYKCVISL-PIFVLIAALQLV 367

RESULT 4
US-08-902-853-1
; Sequence 1, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 394 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LIVRUT04
; CLONE: 2516821
; US-08-902-853-1

```

```

Db      126 RPQLTK-----KFCEASWRFLFYLSFSFGVGLSVLYHESWLMAPVCMCDRYPNQTLPKPSLYW 181
QY      168 FYISQLAYWHPAPPELYFQKTKQDIPROQLVYIGLHLFHITGAYLLVYLNHLGLLLVLHY 227
Db      182 WYLELGFYLSLLIRLPFD-VKKRDKFEQVHHFVAVILMTFFSYSANLLRIGSLVLLLHD 240
QY      228 FVELLSHMCGLFYSDSKYKGIS----LWAIYFILGRVTLIVSVLTGVGFHLAGSONRN 283
Db      241 SSDYLLBCKMWVM--QYQVCDALFLIESFVEFVTRLVLFPTQILYTTYESIS-NRG 297
QY      284 P 284
Db      298 P 298

RESULT      5
US-08-902-853-6
; Sequence 6, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ORIGINAL SOURCE:
; HAPLOTYPE: GenBank
; IMMEDIATE SOURCE:
; LIBRARY: 1675382
; US-08-902-853-6

Query Match 6.7%; Score 128; DB 2; Length 387;
Best Local Similarity 22.0%; Pred. No. 2.4e-06;
Matches 68; Conservative 52; Mismatches 119; Indels 70; Gaps
QY 47 SIYFLTLQHSVAVPAAE-----EQATGSKSYLYTGVKDLATVFFYMLVAIIHATIQE 99

```

```
Query Match      6.7%; Score 128; DB 2; Length 387;
Best Local Similarity 22.0%; Pred. No. 2.4e-06;
Matches 68; Conservative 53; Mismatches 119; Indels 70; Gaps 16;

QY 47 SIVELTLQHSHVAPAAE-----EQATGSKSLYYYGVKDLATVFYFVLMVAIIHTAQE 99
    :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 78 AICFACLLSPSLRPAEPIFLSYKQPDGS---YKGPKDAGCFPIWVIVTFAFRVIMD 134
QY 100 YVLDKI-----NKRMQFTKAKQKNFESGQSVFYFSCIW--GTFILISENCLSDPT 150
Db 135 YVRPVLWNGVRNKVII-----RFGSQG-YSPYYL-CFWFLGLYIYRSNYSWNEE 186
QY 151 LWRKARPHSMFTQMKFFYISOLAYWFHAPPELYFQKTKK---QDIPROLVYVIGLHL--- 204
Db 187 KLPEDYPQYVMSPLFKAYVLIQGFWLOQILVLHLBORADHWQFAHHIVTCAILILSY 246
QY 205 ---FHTGAYLILNHLGLLLL---VLHY--FVELLSHMCGLFYFSDKYQKGLSLWAI 255
Db 247 GFNPLRVGNAILYFDLSYILSGGKMLKYLGFCKICDYLFGIF-----VASWY 296
QY 256 V--FILGRVLTIVS---VLTGVGHLAGSONRNPDALTGNNVLAAKIAVLSSSCTIOAY 310
Db 297 SRHYLESKILRVVVTNAPEIIGFHL-----DVPNGYIFNKPIYIAFILLFTLQL- 347
QY 311 VTWNLTILW 319
Db 348 ---LIYIW 352

RESULT 6
US-08-902-853-3
; Sequence 3, Application US/08902853
; Patent No. 5945330
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HUMAN LONGEVITY-ASSURANCE PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,853
; FILING DATE: Herewith
; CLASSIFICATION: ?
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0345 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 380 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2NOT01
; CLONE: 493014
US-08-902-853-3

Query Match 6.0%; Score 116; DB 2; Length 380;
Best Local Similarity 20.2%; Pred. NO. 5.7e-05;
Matches 56; Conservative 47; Mismatches 104; Indels 70; Gaps 11;
QY 107 KRMQFTKAKQKNFESGQSVFYFSCIWGTFTILISENCLSDPTLIWKARP-HSMMTFQM 165
Db 121 RNRQRPSSLKFKFREASWRFTFYLIATAGMAVIVDKPWFYDMKKVWEGYPIQSTIPSOY 180
QY 166 KEFYISQLAYWFHAPPELYFQKTKQDIPROLVYICGLHFLHITGAYLL---YLNHL--G 219
Db 181 WYMYELSYWLLFS--IASDVKRKDFKEQIIH-----HVATILISFSWFANVIRAG 232
QY 220 LLLLHLHYVELLSHMCGLFYFSDK--YQGISLWAIIVFILGRVLTLIVSVL----- 270
Db 233 TLIMALHDSSDYLLSAAKMFYAGWKNCNNIFIVFAIVFIITRLVILPFWILHCTLVYP 292
QY 271 -----TVGFFHLAGSONRNPDALTGNNVLAAKIAVLSSSCTIOAYVTWNLTILWLO--- 321
Db 293 LELYPAFFGYFFNS-----MMGVQLL-----HIFWAYLILRMAHKF 330
QY 322 ---RWVEDSNIQASCMMKKRSRKRRTENGVGIVETS 355
Db 331 ITGKLVED-----ERSDRETESSEGEAA 355

RESULT 7
US-08-834-467-2
; Sequence 2, Application US/08834467
; Patent No. 6008015
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian
; TITLE OF INVENTION: Glycine Transporter
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,467
; FILING DATE:
; CLASSIFICATION: 45
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-834-467-2

Query Match 5.3%; Score 102; DB 3; Length 687;
Best Local Similarity 19.3%; Pred. No. 0.0058;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;

QY 26 IYSCVGMFFLLGLVFEGETAEASIVFTLQHSVAVPAAEQATGSKSLYYGVKDLATVFF 85
Db 271 LLGCLGVSWL-----VVFCLIRGV-----KSSGKVVYF-----TATPPY 305
QY 86 YMLVAIIIIHATQEVYLDKINRM--QFTKAKONKFNESGQSFVYFFSCINGTFLIS- 142
Db 306 VVLTILFVRGVTLEGAFDGMITYLTPQWDKILEAKVWGDAAQIFVSLACAWGLITMAS 365
QY 143 -----ENCLSDPTLIWKARPHSMMTFQMKFFVYISOLAYWFHAFPELYFOKTKQDIPROL 197
Db 366 YNKFHNCRVDSVII-----SITNCA-----TS 388
QY 198 VYIGLHLFHITGAYLLYLNLHGLLLLVHYFVELLSHMGCLFYFSDKEYQKGIS----- 251
Db 389 YVAGFVIFSILG--FMANHLGVDV-----SRVADHGPGLAFVA--YPEALTLLPISP 436
QY 252 LWAIVF-----ILGRIVTLIV-----SVLTGVPHLAGSONRNDP 285
Db 437 LWSLFFFMILLGLGTQFCLETLVTAIVDEVGNWILQKTKYTVTLGVAVAG----- 489
QY 286 ALTGNNVNLAAKIAVLSSSCTIQAVVTWNLI 316
Db 490 -----FLLGIPLTQAGIYWLLL 507

RESULT 8

US-09-396-177-2
; Sequence 2, Application US/09396177
; Patent No. 6251617
; GENERAL INFORMATION:
; APPLICANT: Albert, Vivian
; TITLE OF INVENTION: Glycine Transporter
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows DEMONSTRATION Version 2.0D
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396,177
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/834,467
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: 314572-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 687 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-396-177-2

Query Match 5.3%; Score 102; DB 4; Length 687;
Best Local Similarity 19.3%; Pred. No. 0.0058;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;

QY 26 IYSCVGMFFLLGLVFEGETAEASIVFTLQHSVAVPAAEQATGSKSLYYGVKDLATVFF 85
Db 271 LLGCLGVSWL-----VVFCLIRGV-----KSSGKVVYF-----TATPPY 305
QY 86 YMLVAIIIIHATQEVYLDKINRM--QFTKAKONKFNESGQSFVYFFSCINGTFLIS- 142
Db 306 VVLTILFVRGVTLEGAFDGMITYLTPQWDKILEAKVWGDAAQIFVSLACAWGLITMAS 365
QY 143 -----ENCLSDPTLIWKARPHSMMTFQMKFFVYISOLAYWFHAFPELYFOKTKQDIPROL 197
Db 366 YNKFHNCRVDSVII-----SITNCA-----TS 388
QY 198 VYIGLHLFHITGAYLLYLNLHGLLLLVHYFVELLSHMGCLFYFSDKEYQKGIS----- 251
Db 389 YVAGFVIFSILG--FMANHLGVDV-----SRVADHGPGLAFVA--YPEALTLLPISP 436
QY 252 LWAIVF-----ILGRIVTLIV-----SVLTGVPHLAGSONRNDP 285
Db 437 LWSLFFFMILLGLGTQFCLETLVTAIVDEVGNWILQKTKYTVTLGVAVAG----- 489
QY 286 ALTGNNVNLAAKIAVLSSSCTIQAVVTWNLI 316
Db 490 -----FLLGIPLTQAGIYWLLL 507

RESULT 9

US-07-757-342D-4
; Sequence 4, Application US/07757342D
; Patent No. 6218509
; GENERAL INFORMATION:
; APPLICANT: IGARASHI, Masao
; MINEGISHI, Takashi
; NAKAMURA, Kazuo
; TITLE OF INVENTION: PROTEIN, DNA AND USE THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
; CUSHMAN
; STREET: 130 Water Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,342D
; FILING DATE: 10-Sep-1991
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BUCKLEY, Linda M.
; REGISTRATION NUMBER: 31003
; REFERENCE/DOCKET NUMBER: 41226
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)523-3400
; TELEFAX: (617)523-6440
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-07-757-342D-4

Query Match 4.9%; Score 94.5; DB 4; Length 696;
Best Local Similarity 20.8%; Pred. No. 0.043;
Matches 75; Conservative 43; Mismatches 103; Indels 139; Gaps 17;

[illegible]

```

RESULT 10
US-08-295-814E-13
; Sequence 13, Application US/08295814E
; Patent No. 5658786
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli E.
; APPLICANT: Borden, Laurence A.
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinshank, Richard L.
; TITLE OF INVENTION: DNA ENCODING TAURINE AND GABA
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/295,814E
; FILING DATE: DECEMBER 19, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 40558-B-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-295-814E-13

```

Query Match 4.8%; Score 93; DB 1; Length 638;
Best Local Similarity 18.9%; Pred. No. 0.057;

Matches	63;	Conservative	42;	Mismatches	89;	Indels	140;	Gaps
Qy	26	IVSCVGMFFLLGLVFECTAEASIVFTLQHSVAVPAEAEQATGSKSLYYGVKDLATVFF	85					
		:::ll:::	:::lll::					
Db	222	LLGCLGYSWV-----VVFLCLIRGV-----KSSGKVYVF-----TATFPY	256					
Qy	86	YMLVAIIIIHATIQEYVLDKINKRQFTKAKQKNES-----CQFSVFFVFFSCIGCTFIL	140					
		:::ll:::	:::lll::					
Db	257	VVLTILFVRGVTTLEGATFGI---MYLTTPKWDKILKAVMGWDAASQIFYSLGCAGWGLIT	313					
Qy	141	IS-----ENCLSDPTLIWKARPHSMMTFQMKFFIYSQLAYWPHAFPELYFQTKKQDIP	194					
		:::ll:::	:::lll::					
Db	314	MASYNKFNHCYRDSVII-----SITNCA-----	337					
Qy	195	RQLVYIGLHLPHTGTAYLLYLNLHGLLLLVHVFVELLSHMCGLFVPSDEKKYOGTIS---	251					
		:::ll:::	:::lll::					
Db	338	-TSYVAGFVIFSI LG---FMANHLGVDP---SRVADHGPGLAEVA---YPEALTLLP	384					
Qy	252	--LWAI VF-----ILGLRLVTLV-----SVLTWGFHLAGSQNR	282					
		lll::ll	:::lll::					
Db	385	ISP LWSLLFFFMILLGLGTFQCLLETFVTAIVDEGVNEWILOKKYTVTLGVAVAG----	440					
Qy	283	NPDALTGNNVYLAAKIAVLSSSCTCIQAYVTWNLI	316					
		:::ll:::	:::lll::					
Db	441	-----FLGLGPI LTSQAGIYWL LL	458					

```

RESULT 11
US-08-240-783B-2
; Sequence 2, Application US/08240783B
; Patent No. 5756348
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Brancheck, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinshank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/240,783B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-240-783B-2

```

Query Match 4.8%; Score 93; DB 1; Length 638;

Best Local Similarity 18.9%; Score 93; DB 3; Length 638;
Matches 63; Conservative 42; Mismatches 89; Indels 140; Gaps 15;

Qy 26 IVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVPAABEQATGSKSLYYGVKDLATVFF 85
Db 222 LIGCLGVSWV-----VFLCLIRGV-----KSSGKVYVF-----TATPPY 256
Qy 86 YMLVAIIIIHATIQEVLDKINKRMQFTKAKONKNES-----GQFSVYFFSCITWGTIL 140
Db 257 VVLTILFVRGVTLEGAFTGI---MYLTPKWDKILEAKVWGDAASQIFSLGCWAGGLIT 313
Qy 141 IS-----ENCLSDPTLIWKARPHSMMTFOMKFFYISOLAYWFHAFPELYFQTKKQDIP 194
Db 314 MASYNKFNHNCYRDSVII-----SITNCA-----337
Qy 195 ROLVYIGLHLHITGAYLLYNHLGLLLLVHLYFVELLSHMCGLFYFSDKEYQKGIS--- 251
Db 338 -TSVYAGFVIFSILG---FMANHLGVDV-----SRVADHGPGLAFVA---YPEALTLLP 384
Qy 252 ---LWAIYV-----ILGRVLTIV-----SVLTGVGFHLAGSQNR 282
Db 385 ISPLWSLLFFFMILLGLGTQFCLETLVTAIVDEVGNEWILQKTYVTLGVAVAG---- 440
Qy 283 NPDALTGNVNVLAAKIAVLSSCTIQAYVTWNL 316
Db 441 -----FLLGIPLTQSAGIYWLL 458

RESULT 12
US-09-084-813-2
; Sequence 2, Application US/09084813
; Patent No. 6127131
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Branche, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinschank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/084.813
; FILING DATE: 26-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-AZ-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-084-813-2

Query Match 4.8%; Score 93; DB 3; Length 638;
Best Local Similarity 18.9%; Pred. No. 0.057;
Matches 63; Conservative 42; Mismatches 89; Indels 140; Gaps 15;

Qy 26 IVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVPAABEQATGSKSLYYGVKDLATVFF 85
Db 222 LIGCLGVSWV-----VFLCLIRGV-----KSSGKVYVF-----TATPPY 256
Qy 86 YMLVAIIIIHATIQEVLDKINKRMQFTKAKONKNES-----GQFSVYFFSCITWGTIL 140
Db 257 VVLTILFVRGVTLEGAFTGI---MYLTPKWDKILEAKVWGDAASQIFSLGCWAGGLIT 313
Qy 141 IS-----ENCLSDPTLIWKARPHSMMTFOMKFFYISOLAYWFHAFPELYFQTKKQDIP 194
Db 314 MASYNKFNHNCYRDSVII-----SITNCA-----337
Qy 195 ROLVYIGLHLHITGAYLLYNHLGLLLLVHLYFVELLSHMCGLFYFSDKEYQKGIS--- 251
Db 338 -TSVYAGFVIFSILG---FMANHLGVDV-----SRVADHGPGLAFVA---YPEALTLLP 384
Qy 252 ---LWAIYV-----ILGRVLTIV-----SVLTGVGFHLAGSQNR 282
Db 385 ISPLWSLLFFFMILLGLGTQFCLETLVTAIVDEVGNEWILQKTYVTLGVAVAG---- 440
Qy 283 NPDALTGNVNVLAAKIAVLSSCTIQAYVTWNL 316
Db 441 -----FLLGIPLTQSAGIYWLL 458

RESULT 13
US-09-343-361-13
; Sequence 13, Application US/09343361
; Patent No. 6225115
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli E. et al
; TITLE OF INVENTION: DNA Encoding Taurine and GABA Transporters and Uses
; FILE REFERENCE: 40558-D
; CURRENT APPLICATION NUMBER: US/09/343.361
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patent In Ver. 2.0 - beta
; SEQ ID NO 13
; LENGTH: 638
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; US-09-343-361-13

Query Match 4.8%; Score 93; DB 4; Length 638;
Best Local Similarity 18.9%; Pred. No. 0.057;
Matches 63; Conservative 42; Mismatches 89; Indels 140; Gaps 15;

Qy 26 IVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVPAABEQATGSKSLYYGVKDLATVFF 85
Db 222 LIGCLGVSWV-----VFLCLIRGV-----KSSGKVYVF-----TATPPY 256
Qy 86 YMLVAIIIIHATIQEVLDKINKRMQFTKAKONKNES-----GQFSVYFFSCITWGTIL 140
Db 257 VVLTILFVRGVTLEGAFTGI---MYLTPKWDKILEAKVWGDAASQIFSLGCWAGGLIT 313
Qy 141 IS-----ENCLSDPTLIWKARPHSMMTFOMKFFYISOLAYWFHAFPELYFQTKKQDIP 194
Db 314 MASYNKFNHNCYRDSVII-----SITNCA-----337
Qy 195 ROLVYIGLHLHITGAYLLYNHLGLLLLVHLYFVELLSHMCGLFYFSDKEYQKGIS--- 251
Db 338 -TSVYAGFVIFSILG---FMANHLGVDV-----SRVADHGPGLAFVA---YPEALTLLP 384
Qy 252 ---LWAIYV-----ILGRVLTIV-----SVLTGVGFHLAGSQNR 282
Db 385 ISPLWSLLFFFMILLGLGTQFCLETLVTAIVDEVGNEWILQKTYVTLGVAVAG---- 440

QY 283 NPDALTGNVNVLAAKIAVLSSSCTIQAYVYTNLI 316
Db 441 -----FLIGIPLTSQAGIYWLIL 458

RESULT 14

US-08-336-031-6
; Sequence 2, Application PC/TUS9209662
; GENERAL INFORMATION:
; APPLICANT: Smith, Kelli
; APPLICANT: Borden, Laurence A.
; APPLICANT: Brantchek, Theresa
; APPLICANT: Hartig, Paul R.
; APPLICANT: Weinshank, Richard L.
; TITLE OF INVENTION: DNA ENCODING A GLYCINE TRANSPORTER AND USES THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/09662
; FILING DATE: 19921112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 1795/39875-A-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 977-9809
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 638 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-09662-2

Query Match 4.8%; Score 93; DB 5; Length 638;
Best Local Similarity 18.9%; Pred. No. 0.057;
Matches 63; Conservative 42; Mismatches 89; Indels 140; Gaps 15;
QY 26 IVSCVGMFFLLGLVFECTAEASIVFLTLQHSVAVPAEAEQATGSKSLYYGVKDLATVFF 85
Db 222 LLGCLGVSWY-----VVFCLIRGV-----KSSGKVYF-----TATFPY 256
QY 86 YMLVAIIHATIQEYVLDKINKRMQFTKAKONKFNES-----GQFSVYFPFCIGCTFIL 140
Db 257 VVLTILFVRGVTLGAPTGI---MYLTPAKDKILCAKVGWDAASQIFSLGCANGGLIT 313
QY 141 IS-----ENCLSDPTLIWKARPHSMMTFQMKFFYISQLAYWPHAPPELYFQKTKQDIP 194
Db 314 MASYNKFNHCYRDSVII-----SITNCA----- 337
QY 195 RQLVYIGLHLFHTAGVALLYNLHGLLLLVHYFVELLSHMCGLFYFSDKYGKGIS--- 251
Db 338 -TSYAGFVIFSILG---FMANHIGVDV-----SRVADHGPGLAEFA---YPEALTLLP 384
QY 252 ---LWAIVF-----ILGLRVLIV-----SVLTGVGFHLGASQNR 282
Db 385 ISPLNSLLFFMLILLGLGTQFCLETLVTAIVDEGVNEWILQKKTIVTLGVAVAG----- 440

QY 283 NPDALTGNVNVLAAKIAVLSSSCTIQAYVYTNLI 316
Db 441 -----FLIGIPLTSQAGIYWLIL 458

RESULT 15

US-08-336-031-6
; Sequence 6, Application US/08336031
; Patent No. 5817782
; GENERAL INFORMATION:
; APPLICANT: Jazwinski, S. M.
; TITLE OF INVENTION: LAG1: A GENE FOR INCREASING THE
; TITLE OF INVENTION: LONGEVITY OF EUKARYOTES
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,031
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,875
; FILING DATE: 03-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 93032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-031-6

Query Match 4.5%; Score 86; DB 2; Length 193;
Best Local Similarity 26.5%; Pred. No. 0.062;
Matches 27; Conservative 18; Mismatches 45; Indels 12; Gaps 3;
QY 33 FPLLGLVFGTAEASIVFLTLQHSVAVP-----AAEQATGSKSLYYGVKDLATVFFYM 87
Db 88 FFILGVY-----SAYFLSGNRNTESNPLHMFVAISYQVDGTDG--YAKGIKDLSFVFFYM 140
QY 88 LVAILIHATIQEYVLDKINKRMQFTKAKONKFNESGQFSVEY 129
Db 141 IFFTFLREFLMDVIRPFTVYLVNVTSEHQRKMLMEQMYAIFY 182

Search completed: September 6, 2002, 17:10:39
Job time: 7973 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:11:51 ; Search time 52.02 Seconds
(without alignments)
681.603 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKKSTKNPPVLSQEFIL.....VGVETSNRVDPPKREKSS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1405.5	73.2	374	2 S21736	translocating chai
2	1392.5	72.5	374	2 S20034	translocating chai
3	494	25.7	371	2 T19417	hypothetical prote
4	487	25.4	373	2 T19419	hypothetical prote
5	144	7.5	411	2 S46800	LAG1 protein - yea
6	128	6.7	390	2 T38012	longevity-assuranc
7	123.5	6.4	418	2 S30134	hypothetical prote
8	116.5	6.1	357	2 T40389	longevity assuranc
9	115	6.0	614	2 B84949	NADH dehydrogenase
10	111.5	5.8	360	2 T27324	hypothetical prote
11	108.5	5.7	476	2 AG1051	probable transpor
12	106.5	5.5	331	2 T20916	hypothetical prote
13	106.5	5.5	459	2 T11411	NADH dehydrogenase
14	106.5	5.5	484	2 A86116	hypothetical prote
15	105.5	5.5	484	2 D65230	hypothetical prote
16	105.5	5.5	484	2 A98275	hypothetical prote
17	102	5.3	638	2 T17912	glycine transporte
18	102	5.3	692	2 T57956	glycine transporte
19	101	5.3	347	2 A53306	dipeptide transpor
20	100.5	5.2	614	2 A59845	Na+/H+ antiporter
21	99.5	5.2	507	2 F37765	hypothetical prote
22	99	5.2	740	2 AB1260	probable integral
23	99	5.2	802	2 S64153	probable membrane
24	98.5	5.1	308	2 H86268	hypothetical prote
25	98.5	5.1	391	2 S07743	ubiquinol--cytochr
26	98.5	5.1	517	2 S21042	cytochrome-c oxida
27	98	5.1	387	2 S07749	hypothetical prote
28	97.5	5.1	506	1 D64048	iron (III) ABC tra
29	96.5	5.0	469	2 D70048	ABC transporter (a

30 96 5.0 1123 2 T48586 hypothetical prote
31 95.5 5.0 534 2 B96642 hypothetical prote
32 95.5 5.0 835 2 T15177 hypothetical prote
33 95.5 5.0 1564 2 T27121 hypothetical prote
34 95 4.9 499 2 H70453 virulence factor M
35 94.5 4.9 489 2 C65147 hypothetical 53.7
36 94.5 4.9 489 2 H91174 probable transport
37 94.5 4.9 489 2 H86020 probable transport
38 94.5 4.9 696 2 A41344 lutropin-choriogon
39 94 4.9 214 2 D83913 hypothetical prote
40 94 4.9 512 2 G90399 amino acid transpo
41 93.5 4.9 394 2 T38922 hypothetical prote
42 93.5 4.9 493 2 S78183 NADH dehydrogenase
43 93.5 4.9 648 2 B84139 ABC transporter (p
44 93 4.8 633 2 I58140 glycine transporte
45 93 4.8 638 2 JH0673 glycine transport

ALIGNMENTS

RESULT 1

S21736

translocating chain-associating membrane protein - dog

C:Species: Canis lupus familiaris (dog)

C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 24-Sep-1999

C:Accession: S21736

R:Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.

Nature 357, 47-52, 1992

A:Title: A protein of the endoplasmic reticulum involved early in polypeptide transla

A:Reference number: S21736; MUID:92244357

A:Accession: S21736

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-374 <GOE>

A:Cross-references: EMBL:X63678; NID:g941; PIDN:CAA45217.1; PID:g942

C:Superfamily: translocating chain-associating membrane protein

C:Keywords: transmembrane protein

Query Match 73.2%; Score 1405.5; DB 2; Length 374;
Best Local Similarity 72.5%; Pred. No. 1.6e-115;
Matches 271; Conservative 40; Mismatches 58; Indels 5; Gaps 2;

Qy 1 MGLRKKSTKNPPVLSQEFILQNHADIVSCVGMFFLLGLVFECTAEASIVFLQLQHSVAVP 60
Db 1 MAIRKSTKSPVLSHFEILQNHADIVSCVAMVLLGLMFEITAKASIIFFVTLQYNVTLUP 60
Qy 61 AAEEQATGSKLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 ATEEQATGTESTLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMHFSKTKHSEN 120
Qy 121 ESGQSFVFFSCINGTIFILISENCLSDPTLIWKARPHSMFTQMKFFYSOLAYWFHAF 180
Db 121 ESGQSAFVFFSCINGTIFILISENIDPTILWRAYPHNLMTQMKFFYIAQLAYWFHAF 180
Qy 181 PELYFQKTKKODIPQLVYIGLHFHITGAYLLYNHLGLLLVLLVHLYFVEFLHSRLDY 240
Db 181 PELYFQKTKKEDIPQLVYIGLHFHITGAYLLYNHLGLVLLVHLYFVEFLHSRLDY 240
Qy 241 FSDEKYQKISLWAVFIILGRVLTIVSVLTGVGFHLAGSQNRNPDLATGNVNLAKIAV 300
Db 241 FSDEKYQKISLWAVFIILGRVLTIVSVLTGVGFHLAGSQNRNPDLATGNVNLAKIAV 300
Qy 301 LSSSTCIQAYVTVNLTILWLQWRVEDSNIQASCMKKK----RSRSKKRKTENV-GVETS 355
Db 301 LASICITQAFMMKMTINFQLRRWRHSHTFQADPVVKKKPTVTGKRSRSGKTENGVTGTVTS 360
Qy 356 NRVDCTPPRKREKSS 369
Db 361 NGADSPRRNRKESKSS 374

```
RESULT 2
S30034
translocating chain-associating membrane protein - human
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C;Accession: S30034
R;Goerlich, D.; Hartmann, E.; Prehn, S.; Rapoport, T.A.
Nature 357, 47-52, 1992
A;Title: A protein of the endoplasmic reticulum involved early in polypeptide translocation
A;Reference number: S21736; MUID:92244357
A;Accession: S30034
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-374 <OE>
A;Cross-references: EMBL:X63679; NID:g37264; PIDN:CAA45218.1; PID:g37265
C;Superfamily: translocating chain-associating membrane protein

Query Match 72.5%; Score 1392.5; DB 2; Length 374;
Best Local Similarity 71.4%; Pred. No. 2.2e-114;
Matches 267; Conservative 44; Mismatches 58; Indels 5; Gaps 2;

QY 1 MGLRKKSTKPNPVLQSQEFILQNHADIIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVP 60
DB 1 MAIRKKSTKPPVLSHSEFVLQNHADIIVSCVAMVFLGLMFETAKASIIIVTLQYNVTL 60
QY 61 AAEEQATGSKSLYYGYGVKDLATVFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 120
DB 61 ATEQATGESVLSYYGIGKDLATVFYMLVAIIHATIQEYVLDKINKRMHFSKTKHKN 120
QY 121 ESGQFSVYFFSCITWGTFLISENCLSDPTLIWKARPHSMFTOMKFFYISQLAYWPHAF 180
DB 121 ESGQLSAFYLPACVWGTFILISENYSIDPTILWRAYPHNLMTFOMKFFYISQLAYWLHAF 180
QY 181 PELYFQTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVVELLSHMCGLFY 240
DB 181 PELYFQTKKEDIPROLVYIGLYLHAGAYLLNLNLGLVLLVHLHFVFEFLHISLFLY 240
QY 241 FSDKEYQKGLSLAIVFILGRVLTVLIVSVLTGPHLAGSNRNPDLTGNVNVLAARIAV 300
DB 241 FSNEKYQKGLSLAVLFLVGLRLTLVLSVLTGPHLAGRAENQKLDFTGPNVLAARIAV 300
QY 301 LSSSCTTQAVYTNWLTITLWLRQWVEDSNIOQSCMKKK-----RSRSKRRKTENG-VGYETS 355
DB 301 LASICVTQAFMMWKFINFQLRRWRHSAFOAPAVKKKPTVTTKGRSSKKGTEGVNGTLTS 360
QY 356 NRVDCCPKRKEKSS 369
DB 361 NVADSPRNKKEKSS 374

RESULT 3
T19417
hypothetical protein C24F3.1a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T19417
R;McMurray, A.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z19122
A;Accession: T19417
A;Status: preliminary; translated from GB/EMBL/DDDB
A;Molecule type: DNA
A;Residues: 1-371 <WTL>
A;Cross-references: EMBL:AL022716; PIDN:CAA18770.1; GSPDB:GN00022; CESP:C24F3.1a
A;Experimental source: clone C24F3
C;Genetics:
A;Gene: CESP:C24F3.1a
A;Map position: 4
A;Introns: 114/3; 158/2; 365/3
C;Superfamily: translocating chain-associating membrane protein

Query Match 25.4%; Score 487; DB 2; Length 373;
Best Local Similarity 32.4%; Pred. No. 5.2e-35;
Matches 120; Conservative 73; Mismatches 143; Indels 34; Gaps 10;

QY 2 GLRKKSTKPNPVLQSQEFILQNHADIIVSCVGMFFLLGLVFEGETAEASIVFLTLQ-----HSV 57
DB 7 GSKASKKPQPPILSHSEFIIQNHGDIIMSCVVMVFIIVGLMFFLTHLSLSLFIAPQINGTYTV 66
QY 58 AVPAAEQATGSKSLYYGYGVKDLATVFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 117
DB 67 AVEQGQEREVHG-----YLSGILDLPALFFYSVCVWIVHVVAVQVEYGLDKISKTKHLSKVSTF 123
QY 118 KFNESGQFSVYFFSCITWGTFLISENCLSD-----PTLIWKARP-----HSMMTFOMKFFYI 170
DB 124 KFGESQHOMFEFTVYS-IAHAFYIVSER-LEDFSEVKSFVSLGYPTEHRVMSAAAYKLYFI 181
QY 171 SOLAYWEHAPPELYFOKTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVE 230
DB 171 SOLAYWEHAPPELYFOKTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVE 230
```

```
Query Match 25.7%; Score 494; DB 2; Length 371;
Best Local Similarity 32.4%; Pred. No. 1.3e-35;
Matches 119; Conservative 74; Mismatches 144; Indels 30; Gaps 9;

QY 2 GLRKKSTKPNPVLQSQEFILQNHADIIVSCVGMFFLLGLVFEGETAEASIVFLTLQ-----HSV 57
DB 7 GSKASKKPQPPILSHSEFIIQNHGDIIMSCVVMVFIIVGLMFFLTHLSLSLFIAPQINGTYTV 66
QY 58 AVPAAEQATGSKSLYYGYGVKDLATVFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 117
DB 67 AVEQGQEREVHG-----YLSGILDLPALFFYSVCVWIVHVVAVQVEYGLDKISKTKHLSKVSTF 123
QY 118 KFNESGQFSVYFFSCITWGTFLISENCLSD-----LSDPTLIWKARP--HSMMTFOMKFFYISQL 173
DB 124 KFGESQHOMFEFTVYS-IAHAFYIVSERLEDFSEVKSWLGYPTTEHRVMSAAAYKLYFIQI 182
QY 174 AYWEHAPPELYFOKTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVELLS 233
DB 183 SYWIHQPEFYLQKLKREINQKSVQAILHIAFISIAFFNFTRVGLALITLEYITOLIF 242
QY 234 HMCGLFYFSEKYGKIGIS-----LWAIIVFILGRVLTVLIVSVLTGPHLAGSNRNPDLT 288
DB 243 HIARFAHVG---RKGLSDPAFLKENGSLVLRGSIILIAVMTWYGLRQAESPFVDISA 299
QY 289 GNVNVLAAKIAVLSSSCTTQAVYTNWLTITLWLRQWVEDSNIOQSCMKKKRS-----RS 341
DB 300 GNFTAVIRLNVLLAVLLQLFLYSFVFMHGRFRESN---AKKEKKKSAASAAVPPKK 356
QY 342 SKRKTEN 348
DB 357 EKKRQDS 363

RESULT 4
T19419
hypothetical protein C24F3.1b - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T19419
R;McMurray, A.
submitted to the EMBL Data Library, April 1998
A;Reference number: Z19122
A;Accession: T19419
A;Status: preliminary; translated from GB/EMBL/DDDB
A;Molecule type: DNA
A;Residues: 1-373 <WTL>
A;Cross-references: EMBL:AL022716; PIDN:CAA18772.1; GSPDB:GN00022; CESP:C24F3.1b
A;Experimental source: clone C24F3
C;Genetics:
A;Gene: CESP:C24F3.1b
A;Map position: 4
A;Introns: 114/3; 158/2; 365/3
C;Superfamily: translocating chain-associating membrane protein

Query Match 25.4%; Score 487; DB 2; Length 373;
Best Local Similarity 32.4%; Pred. No. 5.2e-35;
Matches 120; Conservative 73; Mismatches 143; Indels 34; Gaps 10;

QY 2 GLRKKSTKPNPVLQSQEFILQNHADIIVSCVGMFFLLGLVFEGETAEASIVFLTLQ-----HSV 57
DB 7 GSKASKKPQPPILSHSEFIIQNHGDIIMSCVVMVFIIVGLMFFLTHLSLSLFIAPQINGTYTV 66
QY 58 AVPAAEQATGSKSLYYGYGVKDLATVFYMLVAIIHATIQEYVLDKINKRMQFTKAKON 117
DB 67 AVEQGQEREVHG-----YLSGILDLPALFFYSVCVWIVHVVAVQVEYGLDKISKTKHLSKVSTF 123
QY 118 KFNESGQFSVYFFSCITWGTFLISENCLSD-----PTLIWKARP--HSMMTFOMKFFYI 170
DB 124 KFGESQHOMFEFTVYS-IAHAFYIVSER-LEDFSEVKSFVSLGYPTEHRVMSAAAYKLYFI 181
QY 171 SOLAYWEHAPPELYFOKTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVE 230
DB 171 SOLAYWEHAPPELYFOKTKKODIPROLVYIGLHLFHTIGAYLLYLNHLGLLLVHLHFVE 230
```


A;Cross-references: EMBL:Z28008; NID:g485982; PIDN:CAA81843.1; PID:g485983; MIPS:YKL008C
A;Experimental source: strain S288C
R:Amatruda, J.F.; Gattermeier, D.G.; Cooper, J.A.
submitted to the EMBL Data Library, August 1991
A;Description: yeast capping protein.
A;Reference number: S17016
A;Accession: S17017
A;Molecule type: DNA
A;Residues: 1-149 <AMA>
A;Cross-references: EMBL:X61398; NID:g455515; PIDN:CAA43670.1; PID:g3445
C;Genetics:
A;Map position: 11L
C;Superfamily: hypothetical protein YKL008C
C;Keywords: transmembrane protein

Query Match 6.4%; Score 123.5; DB 2; Length 418;
Best Local Similarity 19.4%; Pred. No. 0.0043;
Matches 59; Conservative 59; Mismatches 117; Indels 69; Gaps 12;

QY 6 KSTKNPPVLSQEF-----ILQNHADI-----VSCVGMFFLLGLVFGTAESIV--FL 51
DB 59 EATNDSDLVKKIWFSPREISYRHWIAPLMLIAVYSAYFTSG-----NTTKTNVLRHV 114
QY 52 TLOHSAVPAEAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQF 111
DB 115 AVSYQI-----GDTNAYGKINDLCPFYFYIMFFTFELFMDVVIRPAIRLHV 164
QY 112 TKAKONFNEGSQSFVY-----FFSCWGTFFILISENCLSDPTLIWKARPHS 159
DB 165 TSKRIKRIEMQMAIFRTVGVSGPGFYCMYHSDLV-----FFNTKAMRTYTPDF 214
QY 160 MMTQFMKFYISQLAYWFAPELYFQTKQDIPROLVYIGL-HLPHITGAYLLYLNHL 218
DB 215 TNPLFKVYLGQAFAQAQACILVLEKPKDHNEITFHHTVLLINSYVFHTKM 274
QY 219 GLLLV-----LHYFVELLSHM-CGLFYFSDKEYOKGISLWAIIVFILGR--LVTLIV 267
DB 275 GLPIYITMDVSDFLSFKTKLYNLDGLAFPS-----FAIFVVAWIYLRHYINLKIWL 327
QY 268 SVLT 271
DB 328 SVLT 331

RESULT 8
T40389
longevity assurance protein homolog SPBC3E7.15c - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C;Accession: T40389; T40499
R:Lyne, M.; Wood, V.; Rajandream, M.A.; Bartell, B.G.; Brown, D.; Churcher, C.M.
submitted to the EMBL Data Library, May 1998
A;Reference number: Z21924
A;Accession: T40389
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <LVN>
A;Cross-references: EMBL:AL023534; PIDN:CAA19018.1; GSPDB:GN00067; SPDB:SPBC3E7.15c
R:Gwilliam, R.; Rajandream, M.A.; Bartell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21933
A;Accession: T40499
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-357 <GWI>
A;Cross-references: EMBL:AL031534; PIDN:CAA20722.1; GSPDB:GN00067; SPDB:SPBC4F6.02c
A;Experimental source: strain 972h-; cosmid c4F6
C;Genetics:
A;Gene: SPBC3E7.15c; SPBC4F6.02c
A;Map position: 2

Query Match 6.1%; Score 116.5; DB 2; Length 357;
Best Local Similarity 21.6%; Pred. No. 0.015;
Matches 50; Conservative 41; Mismatches 121; Indels 19; Gaps 8;
QY 2 GLRKKSTKNPPVLSQEFILQNHAD--IVSCVGMFFLLGLVLF---EGTAESAIVFTLQHS 56
DB 7 GRRRSKSIIVGRAAQNALRSKKEKTWIVPLILLTLLVGVFVNPNGYIKYGI-FL-----S 61
QY 57 VAVPAAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIQEYVLDKINKRMQF-PAK 115
DB 62 YPIEGT-----NPAQYKGRDLIAFLCFYALFTFCREFIMQETIIRIGHFNIRAPAK 115
QY 116 QNKENESQSFVYFFSCWGTFFIL-ISENCLSDPTLIWKARPHSMMTFQMKFFVISOLA 174
DB 116 LRREBQAATCLYFTVMGSLYVAKQTPMWFNTDAFEWEPHYHYVGVSKAFYLIEAA 175
QY 175 YWFHAPPELYFQTK-KODIPROLVYIGLHLFHTGAYLLYLNHLGLLLV 224
DB 176 YWIOQALVLILOLEKPKRDKFELVHHIITLLIGLSYVFHTWIGLAVFI 226

RESULT 9
B84949
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain L [imported] - Buchnera sp. (strain C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: B84949
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
A;Reference number: A84930; MUID:20445173
A;Accession: B84949
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-614 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: nuoL; BUI64
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: NAD; oxidoreductase

Query Match 6.0%; Score 115; DB 2; Length 614;
Best Local Similarity 20.2%; Pred. No. 0.037;
Matches 59; Conservative 52; Mismatches 91; Indels 90; Gaps 14;
QY 80 LATVFF---YMLVAIIHATIQEYVLDKINKRMQFPAKQKNFNEGSQFVSF--YFFSCI 134
DB 1 MSIIFFIILPLIGFLFLSTIQDFIKRYTLNI-----GIFSIFISFFITCF 47
QY 135 WGTILISENCLSDPTLIWKARPHSMMTFQMKF-----FYISQLAYWFAPE 182
DB 48 YGVSLKNNNOVF-TQILWKWL--SINEFKIDFGFDGLSLMLFVITGVGLLIHFSS 104
QY 183 LYFQTKKQDIPROLVYIGLHLFHTGAYL-----LYLNHLGL-----LLLVLYHFVEL 232
DB 105 WYMYKEQS--REFAYTNLFASMSVLVADNLFMLWGLWEGSVCSYLLIGFYTELK 162
QY 233 SHMCGLYFSDKEYOKGISLWAIIVFILGR--VTIIVSVLVGVPHLAGSQNRNPDALTGN 290
DB 163 NNLCAF-----KAFILTRVSDVFLMIGMFLI-YREFNSF----- 196
QY 291 VNVLAAKIATVLSSTCTIOAVVTWNLITLW-----LORWVEDSNI 329
DB 197 ----FQEIKFUSFLNENFYIYDITLTLGLLVIGKSAQLPLQTLWLSDAWV 244

RESULT 10
T27324
hypothetical protein Y6B3B.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27324

R:White, S.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20345

A:Accession: T27324

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-360 <WILL>

A:Cross-references: EMBL:AL032655; PIDN:CAA21723.1; GSPDB:GN00019; CESP:Y6B3B.10

A:Experimental source: clone Y6B3B

C:Genetics:

A:Gene: CESP:Y6B3B.10

A:Map position: 1

A:Introns: 24/3; 58/1; 88/3; 143/1; 205/2; 310/3

Query Match 5.8%; Score 111.5; DB 2; Length 360;
Best Local Similarity 19.7%; Pred. No. 0.041;
Matches 57; Conservative 55; Mismatches 122; Indels 55; Gaps 10;

QY 5 KKSTKNPVLQSEFILQNHADIVSCVGMFFLLGLVFEGETAEASIVFLTLQHSVAVPA--- 61

DB 15 RISTANPIVLA-----GLVFESIPHWFRRYARLRDPDYSPSSMI 53

QY 62 ABEQATGSKSLYYGYKDLATVFFYMLVAIIHA---TIQEVYLDKINKRMQFTKAKQNK 118

DB 54 SDFKKVSLNSSELYTVIIILASIFTLRYVLIQIRLESWTQOHNLYPFAHKVPESFWKLYT 113

QY 119 FNSQSFYVFFSCIWGTGFIILSENCLSDPLIW-----KARPHSMWTFQMKFFYISOL 173

DB 114 YGVTWIFA-FYFHMVCV-----DSHDIFNDPLSMWTEWESGGRP--KMHQVQVIYAVQS 164

QY 174 AYWFAFPELYEOKTKKODIPQLQVYVIGLHLEHITGAYLLYNHL-----GLLLVLHY 227

DB 165 AYIHSIYATLFMDLWRKDSWLMFVH-----HFIALGULFLUSYVDNFTLPGALVFLHD 218

QY 228 FVELLSHMGCLFYFDEKQKQKISLMAIVFILGRVTLIVSVLTGFGHL 276

DB 219 NSDATLEITKLSFLYKKRTNR--QYKYVFLMGNAFAILFAIWIIFRL 265

RESULT 11

AG1051

probable transport protein Sgat sgat [imported] - Salmonella enterica subsp. enterica ser

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AG1051

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AG1051

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD06860.1; PID:gi16505508; GSPDB:GN00176

C:Genetics:

A:Gene: sgat

Query Match 5.7%; Score 108.5; DB 2; Length 476;
Best Local Similarity 22.0%; Pred. No. 0.1;
Matches 73; Conservative 56; Mismatches 136; Indels 67; Gaps 15;

QY 10 NPPVLQSEFILQNHADIVSCVGMFFLL---LGLVFEGETAEASIVFLTLQHSVAVPAEQA 66

DB 29 NAPLL-----LGIIVTCLGYILLKRSVSIIRKTIKTIIGFMLQAGSGILTSTFKP 79

QY 67 TGSKSLYYGVKVD-----LAT-----VFFYMLVAIIHATIQEVYLDKINK 107

DB 80 VVAKMSEVYINGAISDTYASMMATERMGDAYSWGYAVLLALALN--ICVYLLRRITG 137

QY 108 RMOFTKAKQNKENESQFSV-FYFFS-CIWGTFIL-----ISECLSDPTLIWKA 155

DB 138 INTIMLTGHIMFQOAGLIAVSFIYGSWMTTICTAILVSLVWGTISNMWYKPT----- 192

QY 156 RPHSMWTFQMKFF--YISQLAYWFHAFPELYFOKTKKODIPQLQVYVIGLHLEH--ITGAY 211

DB 193 ---QEVTDCCGSIGHQOQFASWI-AKVPAPFLGKKEESVEDLKPGLWNIPHDNIVSTA 248

QY 212 LLYLNHLGLLLLVHYFVELLSHMGCLFYFDEKQKQKISLMAIVFILGRVTLIVSVLT 271

DB 249 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTVYILOTGFSFAVAIFITIGGVFMFAELS 306

QY 272 VGFHLAGSONRNPDALTGNNVLAAKIAVLSS 303

DB 307 EAFN-GISQRLIPGA-----VLAIIDCAIYS 331

RESULT 12

T20916

hypothetical protein F14F8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000

C:Accession: T20916

R:Illyod, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19346

A:Accession: T20916

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-331 <WILL>

A:Cross-references: EMBL:Z92782; PIDN:CAB07188.1; GSPDB:GN00023; CESP:F14F8.3

A:Experimental source: clone F14F8

C:Genetics:

A:Gene: CESP:F14F8.3

A:Map position: 5

A:Introns: 55/3; 95/2; 242/3; 260/3

C:Superfamily: Caenorhabditis elegans hypothetical protein K02H11.4

Query Match 5.5%; Score 106.5; DB 2; Length 331;
Best Local Similarity 24.2%; Pred. No. 0.1;
Matches 63; Conservative 36; Mismatches 80; Indels 81; Gaps 15;

QY 69 SKSLYYGVKDLA-----TVFFYMLVAIIHATIQEVYLDKINKRMQFTKA--- 114

DB 4 SHELKEFTIKDLGIYTSMSFLLSIFLVILFIPIFV---ILNRANRRRDEKTAIYP 60

QY 115 KQKNFESQFSVFFFS--CI-----WGTFILISENCLSDPTLIWKARP 157

DB 61 ITNYFYKS--LCVYIFSVTCIIITVILYGPNAFTGTNIIIFVFFPLIIF-VHV 117

QY 158 HSMWTFQPM---KFFYISQLAYWFHAFPEL--YFOKTKKODIPQLQVYVIGLHLEHITGAYL 212

DB 118 HHLIIFLMAVQRF-----LLVFPFSEPLVTFQCKT-----TNKIYT-LH-----FL 159

QY 213 LYLNLHGLLLLVHYFVELLSHMGCLFYFDEKQKQKISLMAIVFILGRVTLIVSVLT 272

DB 160 FILTHLGLLLDWAYQSLSQIPTLYL-----CYIFLNIQITSAILYI 206

QY 273 GF-----HLAGSQNRNP 284

DB 207 PMVMKIRSFALHASSRTYQP 226

RESULT 13

T11411

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - cat mitochondrion (fragment)

C:Species: mitochondrion Felis silvestris catus (domestic cat)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 12-Nov-1999

```
C:Accession: T11411
R:Lopez, J.V.; Cevario, S.; O'Brien, S.J.
Genomics 33, 229-246, 1996
A:Title: Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial
A:Reference number: 217268; MUID:96301400
A:Accession: T11411
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-459 <LOP>
A:Cross-references: EMBL:U20753; NID:g1098523; PID:g1098533; PIDN:AAC48578.1
A:Experimental source: female adult; isolate FCA-65; lymphocyte; blood
C:Genetics:
A:Genome: mitochondrion
A:Note: ND4
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match          5.5%; Score 106.5; DB 2; Length 459;
Best Local Similarity 21.3%; Pred. No. 0.15;
Matches 51; Conservative 39; Mismatches 82; Indels 67; Gaps 9;

Qy 137 TPILISENCLSDPTLIW-KARPHSMWTQMKEFFYISQLAYWFHAPPELYFQKTKQD--- 192
Db 8 TAMLPMTCLSKPNMIWINSTYLSLISLSYLNQLGHSNLSLFFSDLSAPLLV 67
Qy 193 -----IPQLVYIGLHFLHITGA-----YLLNLNHLGLLLVHYFVELLSHMCGLFY--- 240
Db 68 LTTWLLPLMASQSHLSKEPSRKLYTMTLQLLLIMFTTATELI-----MFYILF 122
Qy 241 -----FSDEKYQKGISLMAIVFILGRVLTIVSVLVGVFHLGASGNRPDAL 287
Db 123 EATLIPTLIITRWGQTERLNAGLYFLFYTLVGLSLPLVALLYI-----QN-----T 170
Qy 288 TGNVNVLAAKI AVLSSCTIQAYVWNLTILW-----LQWVEDSNIOA 331
Db 171 TGTNLFLLIYWAKPIS-----TTWSNIFLWACMAFWKMPYGLHLMLPKRAHVEA 223

RESULT 14
A86116
hypothetical protein sgat [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A86116
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A86116
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-484 <STO>
A:Cross-references: GB:AB005174; NID:g12519184; PIDN:AAG59389.1; GSPDB:GN00145; UWGP:Z58
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: sgat

Query Match          5.5%; Score 106.5; DB 2; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.16;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

Qy 10 NPPVLSQEFILLQNHADIYSCVGMFFL---LGLVFFGTAEASIVELTQLQHSVAVPAAEQA 66
Db 37 NAPLL-----LGI VTCGLYILLRSVSVIIKGTIKTIIGFMLLQAGSGILSTFKP 87
Qy 67 TGSKSLYYGVKVD-----LAT-----VFFYMLVAIIHATIQEYVLDKINK 107
Db 88 VVAKMSEVYINGAISDTYASMMATIDRMGDAYSWGVAVLLALAIN--ICVLLRRITG 145
Qy 108 RMOFTKAKONKFNESGQFSV---FYFFSCIWGTFFL-----ISENCLSDPTLIWKA 155
Db 146 IRTIMLTGHIMFOQAGLIAVTLFIFGYSWMTTIICTAILVSLYWGITSNNMYKPT----- 200
Qy 156 RPHSMWTQMKEFF--YISQLAYWFHAPPELYFQKTKKODIPQLVYIGLHFLH--ITGAY 211
Db 201 ---QEVTDGCGSGIGHQOQFASWI--AYKVAPEFLGKKEVEDLKLPGWLNIFHDNIVSTA 256
Qy 212 LIXLNHLGLLLVHYFVELLSHMCGLFFSDEKYQKGISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTVIYILQTGFSFAVAIFITITQGVRMFVAELS 314
```

```
Db 146 IRTIMLTGHIMFOQAGLIAVTLFIFGYSWMTTIICTAILVSLYWGITSNNMYKPT----- 200
Qy 156 RPHSMWTQMKEFF--YISQLAYWFHAPPELYFQKTKKODIPQLVYIGLHFLH--ITGAY 211
Db 201 ---QEVTDGCGSGIGHQOQFASWI--AYKVAPEFLGKKEVEDLKLPGWLNIFHDNIVSTA 256
Qy 212 LIXLNHLGLLLVHYFVELLSHMCGLFFSDEKYQKGISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTVIYILQTGFSFAVAIFITITQGVRMFVAELS 314
Qy 272 VGFHLGASGNRPDALTCNVNVLAAKIAVLSS 303
Db 315 EAFN-GISORLIPGA-----VLAI DCAAIYS 339

RESULT 15
D65230
hypothetical 52.9 kD protein in aidB-rpsF intergenic region - Escherichia coli (strai
N:Alternate names: hypothetical protein o488
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: D65230; S56418
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D65230
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-484 <BLAT>
A:Cross-references: GB:AB000491; GB:U00096; NID:g2367357; PIDN:AAC77150.1; PID:g23673
A:Experimental source: strain K-12, substrain MG1655
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from
A:Reference number: S56314; MUID:95334362
A:Accession: S56418
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-469, 'AOKKMKNNQNSLNKEF' <BUR>
A:Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97089.1; PID:g537034
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Gene: Yjfs
A:Start codon: GTG

Query Match          5.5%; Score 105.5; DB 2; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.19;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

Qy 10 NPPVLSQEFILLQNHADIYSCVGMFFL---LGLVFFGTAEASIVELTQLQHSVAVPAAEQA 66
Db 37 NAPLL-----LGI VTCGLYILLRSVSVIIKGTIKTIIGFMLLQAGSGILSTFKP 87
Qy 67 TGSKSLYYGVKVD-----LAT-----VFFYMLVAIIHATIQEYVLDKINK 107
Db 88 VVAKMSEVYINGAISDTYASMMATIDRMGDAYSWGVAVLLALAIN--ICVLLRRITG 145
Qy 108 RMOFTKAKONKFNESGQFSV---FYFFSCIWGTFFL-----ISENCLSDPTLIWKA 155
Db 146 IRTIMLTGHIMFOQAGLIAVTLFIFGYSWMTTIICTAILVSLYWGITSNNMYKPT----- 200
Qy 156 RPHSMWTQMKEFF--YISQLAYWFHAPPELYFQKTKKODIPQLVYIGLHFLH--ITGAY 211
Db 201 ---QEVTDGCGSGIGHQOQFASWI--AYKVAPEFLGKKEVEDLKLPGWLNIFHDNIVSTA 256
Qy 212 LIXLNHLGLLLVHYFVELLSHMCGLFFSDEKYQKGISLMAIVFILGRVLTIVSVLT 271
Db 257 IVMTIFFGAIL--LSFGIDTVQAMAGKVHWTVIYILQTGFSFAVAIFITITQGVRMFVAELS 314
```


Qy 272 VGFHLAGSQNRNPDALTGNVNVLAAKIAVLSS 303
|: || | | || | : |
Db 315 EAFN-GISQRLIPGA-----VLAIDCAAIS 339

Search completed: September 6, 2002, 17:11:52
Job time: 6791 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:21:14 ; Search time 26.42 Seconds
(without alignments)
540.784 Million cell updates/sec

Title: US-09-807-470-4

Perfect score: 1920

Sequence: 1 MGLRKKSTKPPVLSQEFIL.....VGVTNRVDCPPKREKSS 369

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1400.5	72.9	373	1	TRAM_CANFA
2	1387.5	72.3	373	1	TRAM_HUMAN
3	1331.5	69.3	358	1	TRAM_BOVIN
4	832.5	43.4	370	1	Y557_HUMAN
5	144	7.5	411	1	LAGL1_YEAST
6	128	6.7	390	1	LAGL1_SCHPO
7	123.5	6.4	418	1	YK88_YEAST
8	116.5	6.1	384	1	YHXF_SCHPO
9	115	6.0	614	1	NUOL_BUCAL
10	106.5	5.5	459	1	NU4M_FELCA
11	105.5	5.5	484	1	SGAT_ECOLI
12	102	5.3	692	1	SGA9_HUMAN
13	99	5.2	802	1	YGN9_YEAST
14	98.5	5.1	391	1	CYB_PARTE
15	98.5	5.1	506	1	HITB_HAETN
16	98.5	5.1	517	1	QOX1_SULAC
17	98	5.1	387	1	YM14_PARTE
18	95.5	5.0	368	1	CYB_TOXGO
19	95	4.9	499	1	MYIN_AQUAE
20	94.5	4.9	489	1	YHIP_ECOLI
21	94.5	4.9	696	1	LSHR_PIG
22	93.5	4.9	394	1	YD2C_SCHPO
23	93	4.8	633	1	S6A9_RAT
24	92.5	4.8	538	1	LSHR_SHEEP
25	91	4.7	638	1	YDAG_BOVIN
26	90.5	4.7	531	1	YDFG_SCHPO
27	90	4.7	1564	1	PDRA_YEAST
28	89	4.6	369	1	Y316_MYCPN
29	89	4.6	633	1	S6A9_MOUSE
30	89	4.6	2841	1	NFI_MOUSE
31	88.5	4.6	457	1	YIRO_YEAST
32	88.5	4.6	496	1	Y048_UREPA
33	88	4.6	461	1	SE12_CAEL

RESULT	1	TRAM_CANFA	STANDARD	PRT	373 AA
AC	Q01685	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	TRAM protein (Translocating chain-associating membrane protein).				
OS	Canis familiaris (Dog)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
OX	NCBI_TaxID=9615;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-26 AND 164-184.				
RC	TISSUE=Kidney;				
RX	MEDLINE=92244357; PubMed=1315422;				
RA	Goerlich D., Hartmann E., Prehn S., Rapoport T.A.;				
RT	"A protein of the endoplasmic reticulum involved early in polypeptide translocation.";				
RL	Nature 357:47-52(1992).				
CC	-1- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF SECRETORY PROTEINS ACROSS THE ER MEMBRANE.				
CC	-1- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.				
CC	-1- SIMILARITY: BELONGS TO THE LASSI FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X63678; CAA45217.1; -				
DR	PIR; S21736; S21736.				
KW	Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.				
FT	INIT_MET 0				
FT	DOMAIN 1 28				CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 29 49				POTENTIAL.
FT	DOMAIN 50 75				LUMENAL (POTENTIAL).
FT	TRANSMEM 76 96				POTENTIAL.
FT	DOMAIN 97 120				CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 121 141				POTENTIAL.
FT	DOMAIN 142 158				LUMENAL (POTENTIAL).
FT	TRANSMEM 159 179				POTENTIAL.
FT	DOMAIN 180 191				CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 192 212				POTENTIAL.
FT	DOMAIN 213 216				LUMENAL (POTENTIAL).
FT	TRANSMEM 217 237				POTENTIAL.
FT	DOMAIN 238 250				CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM 251 271				POTENTIAL.
FT	DOMAIN 272 296				LUMENAL (POTENTIAL).
FT	TRANSMEM 297 317				POTENTIAL.
FT	DOMAIN 318 373				CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD 55				N-LINKED (GLCNAC...) (PROBABLE).
SQ	SEQUENCE 373 AA; 43029 MW; 1D85808E1D80E835 CRC64;				

Q22549 caenorhabdi
Q20076 caenorhabdi
P75134 mycoplasma
O67472 aquifex aeo
P40886 saccharomyc
Q04182 saccharomyc
P75463 mycoplasma
O84068 chlamydia t
Q9hcx4 homo sapien
P97526 rattus norv
P21359 homo sapien
P81318 methanococc

ALIGNMENTS

```
Query Match 72.9%; Score 1400.5; DB 1; Length 373;
Best Local Similarity 72.6%; Pred. No. 7.5e-108;
Matches 270; Conservative 40; Mismatches 57; Indels 5; Gaps 2;

QY 3 LRKSTKNPPVLSOEFTLQNHADIVSCVGMFFLLGLVFEQTAEASIVFLQLQHSVAVPAA 62
DB 2 IRKSTKSPVLSHEFTLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNVTLPAT 61
QY 63 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 122
DB 62 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 121
QY 123 GQSFVFFFCINGTFTLISENCISDPTLIWKARPHSMFTQMKFFYISOLAYWFHAFPE 182
DB 122 GQSFVFFFCINGTFTLISENCISDPTLIWKARPHSMFTQMKFFYISOLAYWFHAFPE 181
QY 183 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 242
DB 182 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 241
QY 243 DEKYQKISLWAVFLGRVTLVSVLTGFGHLAGSONRNPDLTGNVNVLAIAVL 302
DB 242 DEKYQKISLWAVFLGRVTLVSVLTGFGHLAGSONRNPDLTGNVNVLAIAVL 301
QY 303 SSCITQAVTWNLTLMQWRVEDSNIOASCMKKK----RSRSKKRTENG-VGVTSNR 357
DB 302 SICITQAFMMKWFNFOLRRWREHSTFQAPVVKKKPTVTKGRSRKGTENGVTGTSNG 361
QY 358 VDCPPKREKSS 369
DB 362 ADSPRNKKKSS 373

RESULT 2
TRAM_HUMAN ID TRAM_HUMAN STANDARD; PRT; 373 AA.
AC Q15629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TRAM protein (translocating chain-associating membrane protein).
GN TRAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92244357; PubMed=1315422;
RA Goerlich D., Hartmann E., Prehn S., Rapoport T.A.;
RT "A protein of the endoplasmic reticulum involved early in polypeptide translocation."
RL Nature 357:47-52(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Kidney;
RA Strausberg R.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLOCATION OF SECRETORY PROTEINS ACROSS THE ER MEMBRANE.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum membrane.
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC
```

```
DR EMBL; X63679; CAA45218.1; -.
DR EMBL; BC000687; AA00687.1; -.
DR MIM; 605190; -.
KW Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
FT INIT_MET 0 0
FT DOMAIN 1 28
FT TRANSMEM 29 49
FT DOMAIN 50 75
FT TRANSMEM 76 96
FT DOMAIN 97 120
FT TRANSMEM 121 141
FT TRANSMEM 142 158
FT TRANSMEM 159 179
FT DOMAIN 180 191
FT TRANSMEM 192 212
FT DOMAIN 213 216
FT TRANSMEM 217 237
FT TRANSMEM 238 250
FT TRANSMEM 251 271
FT TRANSMEM 272 296
FT TRANSMEM 297 317
FT TRANSMEM 318 373
FT CARBOHYD 55 55
FT SEQUENCE 373 AA; 42940 MW; C220949AF4EFEDD0 CRC64;

Query Match 72.3%; Score 1387.5; DB 1; Length 373;
Best Local Similarity 71.5%; Pred. No. 8.8e-107;
Matches 266; Conservative 44; Mismatches 57; Indels 5; Gaps 2;

QY 3 LRKSTKNPPVLSOEFTLQNHADIVSCVGMFFLLGLVFEQTAEASIVFLQLQHSVAVPAA 62
DB 2 IRKSTKSPVLSHEFTLQNHADIVSCVAMVFLGLMFEITAKASIIFFVTLQYNVTLPAT 61
QY 63 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 122
DB 62 EQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKFNES 121
QY 123 GQSFVFFFCINGTFTLISENCISDPTLIWKARPHSMFTQMKFFYISOLAYWFHAFPE 182
DB 122 GQSFVFFFCINGTFTLISENCISDPTLIWKARPHSMFTQMKFFYISOLAYWFHAFPE 181
QY 183 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 242
DB 182 LYFQTKKQDIPROLYVIGLHFLHITGAYLLYNHLGLLLVLLHYFVELLSHMCGLFYFS 241
QY 243 DEKYQKISLWAVFLGRVTLVSVLTGFGHLAGSONRNPDLTGNVNVLAIAVL 302
DB 242 DEKYQKISLWAVFLGRVTLVSVLTGFGHLAGSONRNPDLTGNVNVLAIAVL 301
QY 303 SSCITQAVTWNLTLMQWRVEDSNIOASCMKKK----RSRSKKRTENG-VGVTSNR 357
DB 302 SICITQAFMMKWFNFOLRRWREHSTFQAPVVKKKPTVTKGRSRKGTENGVTGTSNV 361
QY 358 VDCPPKREKSS 369
DB 362 ADSPRNKKKSS 373

RESULT 3
TRAM_BOVIN ID TRAM_BOVIN STANDARD; PRT; 358 AA.
AC Q9GKZ4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TRAM protein (translocating chain-associating membrane protein) (Fragment).
GN TRAM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
OC
```

```
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Clark T.G., Morris J., Akamatsu M., McGraw R.A., Ivarie R.D.;
RT "Cloning and sequence analysis of a bovine tram cDNA.";
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STIMULATORY OR REQUIRED FOR THE TRANSLLOCATION OF
CC SECRETORY PROTEINS ACROSS THE ER MEMBRANE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM MEMBRANE (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U19578; AAG10391.1; -.
KW Endoplasmic reticulum; Transmembrane; Glycoprotein; Translocation.
FT NON-TER 1 1
FT DOMAIN <1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL.
FT DOMAIN 35 60 LUMENAL (POTENTIAL).
FT TRANSMEM 61 81 POTENTIAL.
FT DOMAIN 82 105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 106 126 POTENTIAL.
FT DOMAIN 127 143 LUMENAL (POTENTIAL).
FT TRANSMEM 144 164 POTENTIAL.
FT DOMAIN 165 176 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 177 197 POTENTIAL.
FT DOMAIN 198 201 LUMENAL (POTENTIAL).
FT TRANSMEM 202 222 POTENTIAL.
FT DOMAIN 223 235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 236 256 POTENTIAL.
FT DOMAIN 257 281 LUMENAL (POTENTIAL).
FT TRANSMEM 282 302 POTENTIAL.
FT DOMAIN 303 358 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC... (PROBABLE).
SQ SEQUENCE 358 AA; 41403 MW; 25CF9930C4CDDA15 CRC64;

Query Match 69.3%; Score 1331.5; DB 1; Length 358;
Best Local Similarity 71.8%; Pred. No. 3.3e-102;
Matches 257; Conservative 37; Mismatches 59; Indels 5; Gaps 2;

QY 17 EFILQNHADIVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVPAABEQATGSKSLYYG 76
DB 1 EFVLQNHADIVSCVAMVFLGLMFEITAKVSIIFVTLLQNVNLTLPATEEQATESAFLLYYG 60
QY 77 VKDLATVFFVLMVAIIHATIQEYVLDKINKRQFTKAKONKNESQPSVFFFSICWG 136
DB 61 IKDLATVFFVLMVAIIHATIQEYVLDKINKRHFHSKTHSKFNESQGLSAFLFSICWG 120
QY 137 TFLISENCLSDPTLWKARPHSMFTOMKFFVYSOLAYWFHAPPELYFOKTKKQDIPRO 196
DB 121 TFLISENYSIDPTLWRAVPHNLMTQMAFFVYSOLAYWFHAPPELYFOKTKKQDIPRO 180
QY 197 LVYIGLHLFHITGAYLLYLNLHGLLLLVHYFVELLSHMCGLFFSDEKQKGLSLMAIV 256
DB 181 LVYIGLYLPHIAGAILNLNLHGLVLLVLYHVEFFLHISRLFEYFDEKQKGLSLMAVL 240
QY 257 FILGRVLTVLSVTVGFLHAGSONRNPDALGTGNVNLAAKIAVLSSSCTIQAYVTWNLI 316
DB 241 FVLGRLLTLVSVTVGFLARAENQKLDFTSGNFNVAVRIAVLASICITQAFMMWKEI 300
QY 317 TLWLQWRVDSNTQASCKKK-----RSRSKKRTKENGCV-GVETSNRVDCPPKREKSS 369
DB 301 NFQLRRWRHSATQAPAVKKKPPVTGKRSXXKGTENGVTGTVTNGADSPNRKREKSS 358
```

```
RESULT 4
Y557 HUMAN STANDARD; PRT; 370 AA.
AC Q15035;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0057.
GN KIAA0057.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Tracey A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D31762; BAA06540.1; -.
DB EMBL; AL049611; CAB71119.1; -.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 23 43 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 251 271 POTENTIAL.
FT TRANSMEM 288 308 POTENTIAL.
SQ SEQUENCE 370 AA; 43327 MW; 9B5183F1A3D45366 CRC64;

Query Match 43.4%; Score 832.5; DB 1; Length 370;
Best Local Similarity 45.1%; Pred. No. 3.1e-61;
Matches 171; Conservative 72; Mismatches 115; Indels 21; Gaps 7;

QY 1 MGLRKKSTKPNPVLVSQEFILQNHADIVSCVGMFFLLGLVFEPTAEASIVFLTLQHSVAVP 60
DB 1 MAPRRR-TKSYPLSFQEFVHNHADIGFCLVLCVGLMFEVTAFTKFLFILPOYNISVP 59
QY 61 AABEQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRQFTKAKONKNFN 120
DB 60 TADSETVH----YHYGPKDLVTILFYITILHAVVQEVYLDKISKRLHLSKVHKSFN 115
QY 121 ESQFSVFFFSICWGTGFLISENCLSDPTLWKARPHSMFTOMKFFVYSOLAYWFHAF 180
DB 116 ESQGVVFFHFTSVTWCFYVVVVVTEGYLTNPRSLWEDYVILDKISKRLHLSKVHKSFN 175
QY 181 PELYFOKTKKQDIPROQLVYIGLHLFHITGAYLLYLNLHGLLLLVHYFVELLSHMCGLFY 240
DB 176 PELYFOKVRKEEIPROQLVYIGLHLFHITGAYLLYLNLHGLLLLVHYFVELLSHMCGLFY 235
QY 241 FSDKQYKGLSLMAIVFILGRVLTVLIVSVTVGFLHAGSONRNPDALGTGNVNLAAKIAV 300
```

```
Db 236 FADENNEKLSAWA VFGVTRFILTLAVLAIGFGLARMENQAFDPEKGFNTLFCLVCV 295
QY 301 LSSCTIQAYVWNLTLWLQWRVSDSNIOASCKKR-----SRSSKRT---ENG 349
Db 296 LLLVCAAQAWLWRETHSOLRWETWNEQSA---KRRVPATPRLPARLIKRESGYHENG 352
QY 350 VGVETSNRVDCPPKRRKS 368
Db 353 V-VKAENGTS-PRTKLKS 369

RESULT 5
LAG1 YEAST
ID LAG1 YEAST STANDARD; PRT; 411 AA.
AC P38703;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR YHL003C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-X2180-1A;
RX MEDLINE-94253121; PubMed-8195187;
RA D'Mello N.P., Childress A.M., Franklin D.S., Kale S.P.,
RA Pinwasdi C., Jawinski S.M.,
RT "Cloning and characterization of LAG1, a longevity-assurance gene in
RT yeast.";
RL J. Biol. Chem. 269:15451-15459(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-5288C / AB972;
RX MEDLINE-94378003; PubMed-8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favella A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Lateille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasikis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
CC -!- FUNCTION: INVOLVED IN THE AGING PROCESS. DELETION OF LAG1 RESULTS
CC IN A PRONOUNCED INCREASE (APPROXIMATELY 50%) IN MEAN AND IN
CC MAXIMUM LIFE SPAN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U08133; AAA21579.1; -.
CC DR EMBL; U10555; AAB68429.1; -.
CC DR PIR; S46800; S46800.
CC DR SGD; S0000595; LAG1.
KW Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 212 232 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
```

```
FT CONFLICT 173 174 ML -> IV (IN REF. 2).
FT CONFLICT 220 220 F -> C (IN REF. 2).
FT CONFLICT 301 411 VFTPEVGLVFEVFWLYLRHVNVNIRILWSVLTFRHEGVNVL
FT NPATQYKWCISLPVFWLIAALQVLYWFLIRILYRL
FT IWQGLQKDRSDSDSEAEENESKEKE -> TEISGIWE
FT KOEIDSDNDPRTALSPNETSKQVKKPDLVLVLPENRNAL
FT LEAIKSRVPTAITDIDSEPSLVYPIFGNDDSLRSVNFLL
FT GVLARAGQGLQNLARNNEK (IN REF. 1).
SQ SEQUENCE 411 AA; 48454 MW; 91676D56AC053F3C CRC64;

Query Match 7.5%; Score 144; DB 1; Length 411;
Best Local Similarity 22.6%; Pred. No. 0.00011;
Matches 65; Conservative 49; Mismatches 112; Indels 62; Gaps 10;

QY 33 FFLGLVPEGTAEASIVFLTLOHSAVNP-----AAEEQATGSKSLYYGVKDLATVFFFM 87
Db 88 FFILVCVY-----SAYFLSGNRSTESNPLHMFVAISYQVDGDS-YAKIKDLSFVFFYM 140
QY 88 LVAIIHATIQEYVLVDKINKRMQFTKAKONKFNESGQSVFYF-FSCIWGTFLI-SENC 145
Db 141 IFFTLREFLMDVIRPFTVYLVNTSEHRQKRMLEQMYAIFYGVSGPGLYIMYHSDLW 200
QY 146 LSDPTLINKARPHSMFTQMKFFYISQLAYWFAHPPELYFQTKKODIPRQLVYIGL-HL 204
Db 201 LFKTKPMVRYTPVITNPFELFKFYLGQAFAWAQACVLVLQLEKPKDYKELVFHHIVTL 260
QY 205 FHITGAYLLYLNHLGLLLV-----LHYFVELLS-HMCGLFY----- 240
Db 261 LLIISSYVFHFTKMGAIYITMDVSDFFLSLSKTLNLSVTFPFGVLFVFFWIYLRHV 320
QY 241 -----FSDEKYQKIGSLWAIVFILGRVLTLI 266
Db 321 VNIRILWSVLTFRHEGVNVLNFAQTQYKWCISL-PIFVFLAALQLV 367

RESULT 6
LAG1 SCHPO
ID LAG1 SCHPO STANDARD; PRT; 390 AA.
AC P78970; Q13860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Longevity-assurance protein 1 (Longevity assurance factor 1).
GN LAG1 OR SPAC1A6.09C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Chanda E.R., Lingner C., Ko Z., Young P.G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-972;
RA Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V., Wood V.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE AGING PROCESS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U76608; AAB19113.1; -.
CC DR EMBL; Z99258; CAB16359.1; -.

```

KW Transmembrane.
FT DOMAIN 13 16 POLY-SER.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 336 356 POTENTIAL.
FT CONFLICT 379 390 EDEASSTNEDK -> GRGGGFENE (IN REF. 1).
SQ SEQUENCE 390 AA; 45668 MW; DC00FB5C2D2F22CC CRC64;

Query Match 6.7%; Score 128; DB 1; Length 390;
Best Local Similarity 22.0%; Pred. No. 0.0022;
Matches 68; Conservative 52; Mismatches 119; Indels 70; Gaps 16;

QY 47 SIVFLQLQSHVAVPAE-----EQATGSKSLYYGVKDLATVFFYMLVAIIHATIOE 99
DB 78 AICFACLLSPSLRPAEPIFTLSYKQPDGS---YKGPKDACPPIFWIVFTAFRVIMVD 134
QY 100 YVLDKI-----NKRMQTKAKQNFESGQFVSFFSCIW--GTFILISENCLSDPT 150
DB 135 YVPRPVLNMGVNRKVVII-----RFCQSG-YSFYFYL-CFVFLGLIYRSNYSNNEE 186
QY 151 LIWKARPHSMMTQMKFFYISQLAYWFHAPPELYFQKTK---QDIPROLVYIGLHL--- 204
DB 187 KLEFEDYQYMYMPLFRAYLYLQGLFWLQILYHLQRRADHWQMFHAHIVTCALILSY 246
QY 205 ---FHTGAYLYNLHGLLLL-----VLHY--FVELLSHMCGLFYFSDKQYKGISLWAI 255
DB 247 GFNFLRVGNAILYFDLSYLLGGKMLYLGKICDYLFGIF-----VASWVY 296
QY 256 V--FILGRVLTIVS---VLTGVFHLAGSNRPDALTGNVNVLA KIAVLSSCTIQAY 310
DB 297 SRHLYFSKILRVVVTNAPETIIGFHL-----DVPNGYIFNKPIYIAFIILLETQL- 347
QY 311 VTWNLTWL 319
DB 348 ----LIIYI 352

RESULT 7
YKAB YEAST
ID YKAB YEAST STANDARD; PRT; 418 AA.
AC P28496;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 49.0 kDa protein in UFD4-CAP1 intergenic region.
GN YKL008C OR YKL156.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93255906; PubMed=8488728;
RA Boyer J., Pascolo S., Richard G.F., Dujon B.;
RT "Sequence of a 7.8 kb segment on the left arm of yeast chromosome XI
RT reveals four open reading frames, including the CAP1 gene, an intron-
RT containing gene and a gene encoding a homolog to the mammalian UOG-1
RT gene.";
RL Yeast 9:279-287(1993).
RN [2]
RP SEQUENCE OF 1-149 FROM N.A.
RX MEDLINE=93077675; PubMed=1447293;
RA Amatruda J.F., Gattermeir D.J., Karpova T.S., Cooper J.A.;
RT "Effects of null mutations and overexpression of capping protein on
RT morphogenesis, actin distribution and polarized secretion in yeast.";
RL J. Cell Biol. 119:1151-1162(1992).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE LASS1 FAMILY.

CC -----SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X61398; CAA43670.1; -;
DR EMBL; S59773; AAC60549.1; -;
DR EMBL; Z28008; CAA81843.1; -;
DR PIR; S30134; S30134.
DR SGD; S0001491; YKL008C.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 173 193 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
SQ SEQUENCE 418 AA; 48992 MW; 7691BA623AC0460A CRC64;

Query Match 6.4%; Score 123.5; DB 1; Length 418;
Best Local Similarity 19.4%; Pred. No. 0.0056;
Matches 59; Conservative 59; Mismatches 117; Indels 69; Gaps 12;

QY 6 KSTKNPVLISQEF-----ILQNHADI-----VSCVGMFFLGLVFEPTAEASIV--FL 51
DB 59 EATKNDSDLVKVIKWFISFREISYRHWIAPLMILIAVYSAYFTSG---NTTKTNVLRHV 114
QY 52 TLOHSAVVAFAAEQATGSKSLYYGVKDLATVFFYMLVAIIHATIOEYVLDKINRQMF 111
DB 115 AVSYQI-----GDTNAYGKGINDLFCFYIMFFTFLEFLMDVIVIRFAIRLHV 164
QY 112 TRAKONKFNESGQFVSFY-----FFSCIWGTFLISENCLSDPTLIWKARPHS 159
DB 165 TSKHRIKRIEMQYAIPTGVSQPGFIYCMYHSDLW-----FFNTKAMRYTPDF 214
QY 160 MMTFQMKFFYISQLAYWFHAPPELYFQKTKQDIPROLVYIGL-HLFHITGAYLLYNHL 218
DB 215 TNPFLKVFYVGOAFWAQOACILVLQLEKPRKDHNLTFHHIVTLLLIWSSVVFHTKM 274
QY 219 GLLLLV-----LHYFVELLSHM-CGLFYFSDKQYKGISLWAIIFILGR---LTVLIV 267
DB 275 GLPIYITMDVSDFLSFKTLNLYDSGLAFFS-----FAIFVAVIYLRHYINLKILW 327
QY 268 SVLT 271
DB 328 SVLT 331

RESULT 8
YHXF SCHPO
ID YHXF SCHPO STANDARD; PRT; 384 AA.
AC O59735;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 42.4 kDa protein C3E17.15c in chromosome II.
GN SPBC3E7.15C OR SPBC4F6.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Lyne M., Wood V., Rajandream M.A., Barrell B.G., Brown D.,
RA Churcher C.M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.

```

RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=972;
RA  Gwilliam R., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
RL  Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -!- SIMILARITY: BELONGS TO THE LASSI FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AL023534; CAAL9018.2; -
DR  EMBL; AL031534; CAAT0722.2; -
KW  Hypothetical protein; Transmembrane.
FT  TRANSMEM 59 79 POTENTIAL.
FT  TRANSMEM 104 124 POTENTIAL.
FT  TRANSMEM 155 175 POTENTIAL.
FT  TRANSMEM 195 215 POTENTIAL.
FT  TRANSMEM 233 253 POTENTIAL.
FT  TRANSMEM 270 290 POTENTIAL.
FT  TRANSMEM 329 349 POTENTIAL.
SQ  SEQUENCE 384 AA; 45335 MW; 82FCF8EA6638849A CRC64;

Query Match          6.18; Score 116.5; DB 1; Length 384;
Best Local Similarity 21.64; Pred. No. 0.019;
Matches 50; Conservative 41; Mismatches 121; Indels 19; Gaps 8;

Qy  - 2 GLRKKSTKNPPVLSQEFILQNHAD--IVSCVGMFLLGLVF---EGTAEASIVFTLQHS 56
Db  34 GRRRSKSIIVGRAQNAVLRSEKRTWIVPLILLVLLVGVFWPNNGYIKYGI-FL----S 88

Qy  57 VAVPAEEQATGSKSLXYGVKDLATVFFYMLVAIIHATIOEYVLDKINKRMQF-TRAK 115
Db  89 YPIPT-----NPAQYKGRDLIAFLYALFTFCFEITMQEIIARIGRHFNIRAPAK 142

Qy  116 QNKFNESQSFVFFSCINGTFFIL-ISENCLSDPTLLWKARPHSMFTQMKFFYISOLA 174
Db  143 LRRFEQAYTCLFYFVMSGLYVMSQKPMFMFFNTDAFWEEYHPHYHVGSPKAFYLIEAA 202

Qy  175 YWFHAFPELYFOKTK-KQDIPRLQVYIGLHLFHTGAYLLVNLHLLLV 224
Db  203 YWIOQALVLLIQLEKPRKDFELVHVHIIITLLIGLSYFFHTWIGLAVFI 253

RESULT 9
NUOL_BUCAI          STANDARD;          PRT;          614 AA.
AC  P57262;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DE  NADH dehydrogenase I chain L (EC 1.6.5.3) (NADH-ubiquinone
DE  oxidoreductase chain L).
GN  NUOL OR BU164.
OS  Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS  symbiotic bacterium).
OC  Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX  NCBI_TaxID=118099;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-TOKYO 1998;
RX  MEDLINE=20445173; PubMed=10930377;
RA  Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT  "Genome sequence of the endocellular bacterial symbiont of aphids
RT  Buchnera sp. APS."
RL  Nature 407:81-86(2000).
CC  -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
```

```

CC  -!- SUBUNIT: COMPOSED OF 13 DIFFERENT SUBUNITS. SUBUNITS NUOA, H, J,
CC  K, L, M, N CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX (BY
CC  SIMILARITY).
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC  -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC  OF CHLOROPLASTS OR MITOCHONDRIA.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AP001118; BAB12882.1; -
DR  InterPro; IPR001750; Oxidored_q1
DR  InterPro; IPR001516; Oxidored_q1_N.
DR  Pfam; PF00361; oxidored_q1; 1.
DR  Pfam; PF00662; oxidored_q1_N; 1.
KW  Oxidoreductase; NAD; Ubiquinone; Transmembrane; Complete proteome.
FT  TRANSMEM 1 21 POTENTIAL.
FT  TRANSMEM 33 53 POTENTIAL.
FT  TRANSMEM 79 99 POTENTIAL.
FT  TRANSMEM 136 156 POTENTIAL.
FT  TRANSMEM 168 188 POTENTIAL.
FT  TRANSMEM 207 227 POTENTIAL.
FT  TRANSMEM 247 267 POTENTIAL.
FT  TRANSMEM 271 291 POTENTIAL.
FT  TRANSMEM 327 347 POTENTIAL.
FT  TRANSMEM 372 392 POTENTIAL.
FT  TRANSMEM 410 430 POTENTIAL.
FT  TRANSMEM 455 475 POTENTIAL.
FT  TRANSMEM 492 512 POTENTIAL.
FT  TRANSMEM 533 553 POTENTIAL.
FT  TRANSMEM 593 613 POTENTIAL.
SQ  SEQUENCE 614 AA; 70881 MW; 18CCC2DFC4FE27E0 CRC64;

Query Match          6.0%; Score 115; DB 1; Length 614;
Best Local Similarity 20.28; Pred. No. 0.043;
Matches 59; Conservative 52; Mismatches 91; Indels 90; Gaps 14;

Qy  80 LATVFF---YMLVAIIHATIOEYVLDKINKRMQFTKAKQKNFESQFSVF--YFFSCI 134
Db  1 MSIIFFILPLIGFLSTIODTFKRYTNI-----GIFSIFISFPTCF 47

Qy  135 WGTFILISENCLSDPTLLWKARPHSMFTQMKF-----FYISQLAYWHPAPPE 182
Db  48 YGVSIKNNNQVF-TQILMKWL--SINEFKIDFGFFLDGLSLMLFVITGVGLLIHFSS 104

Qy  183 LYFKTKKODIPRLQVYIGLHLFHTGAYL-----LYLNHLGL-----LLVLVHYFVELL 232
Db  105 WMYRKEQS--RFFAYTNLFIASMSVLVLADNLFMLGWEGVSVCSYLLIGFYITELK 162

Qy  233 SHMGLFYFSDKYQKIGISLWAIIFILGR--VTILIVSLVTVGFHLAGSQNRNPDALGN 290
Db  163 NNLCAF-----KAFILTRVSDVFLMIGMELI-YREFNSFN----- 196

Qy  291 VNVLAATVLSSTCTQAYVTWNLIILW-----LQRWVEDSNI 329
Db  197 ----FQEKFLSSFLNVENFYLDYITLILLGVIGKSAQLPLQTLSDAMV 244

RESULT 10
NU4M_FELCA
ID  NU4M_FELCA          STANDARD;          PRT;          459 AA.
AC  P48916;
DT  01-FEB-1996 (Rel. 33, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DT  01-NOV-1997 (Rel. 35, Last annotation update)
DE  NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).
GN  MTND4 OR ND4.
```



```
OS Felis silvestris catus (Cat).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=96301400; PubMed=8660972;
RA Lopez J.V., Cevario S., O'Brien S.J.;
RT "Complete nucleotide sequences of the domestic cat (Felis catus) in
RT mitochondrial genome and a transposed mtDNA tandem repeat (Numt) in
RT the nuclear genome.";
RL Genomics 33:229-246(1996).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U20753; AAC48578.1; -.
DR InterPro; IPR003918; NADH_oxdrdctse4.
DR InterPro; IPR001750; Oxidored_q1.
DR InterPro; IPR000260; Oxidored_q5_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
DR PRINTS; PR01437; NUOXDRDTASE4.
KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
SQ SEQUENCE 459 AA; 51497 MW; C1FA567606F1BFB2 CRC64;

Query Match 5.5%; Score 106.5; DB 1; Length 459;
Best Local Similarity 21.3%; Pred. No. 0.15;
Matches 51; Conservative 39; Mismatches 82; Indels 67; Gaps 9;

QY 137 TELISENCSDPTLW-KARPSMFTQKFFYISQLAWFHAFPELYFQKTKQD--- 192
Db 8 TAMDPTKUSKPNMIWINSTTYSLSISLSISLNLGSHSNFSLFFSDSLAPLV 67
QY 193 -----IPRQVYIGLHLFHTGA-----YLYLNHLGLLLVLHYFVLLSHMGLFY--- 240
Db 68 LTTWLLPLMLASQSHLSKETPSRKLYITMLTLLQLLLIMTFTATELI-----MEYILF 122
QY 241 -----FSDEKYQKGISLNAIVPILGRVLTVLSVLTGVFHLAGSONRNPDAL 287
Db 123 EATLIPTLIITRWGQDTERLNAGLYFLFTVLGSLPLVALLYI-----QN-----T 170
QY 288 TGNVNVLAAKIALLSSCTIQAVVTWNLTW-----LQRWVEDSNIOA 331
Db 171 TGTINFLIIQYWKPIIS-----TTSNIFLWLCMAFMVKMPLYLGLHLWLPKRAHVEA 223

RESULT 11
SGAT_ECOLI
ID SGAT_ECOLI STANDARD; PRT; 484 AA.
AC P39301;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative transport protein sgat.
GN SGAT OR B4193.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
```

```
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP DISCUSSION OF SEQUENCE.
RA Reizer J., Charbit A., Reizer A., Saier M.H. Jr.;
RT "Novel phosphotransferases system genes revealed by bacterial genome
RT analysis: operons encoding homologues of sugar-specific permease
RT domains of the phosphotransferase system and pentose catabolic
RT enzymes.";
RL Genome Sci. Technol. 1:53-75(1996).
CC -!- FUNCTION: COULD ACT AS THE TRANSPORT PROTEIN FOR THE UNKNOWN
CC PENTITOL SUBSTRATE OF THE SGA OPERON.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: TO M.PNEUMONIAE SGAT HOMOLOG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14003; AAA97089.1; -.
DR EMBL; AE000491; AAC77150.1; -.
DR EcoGene; EG12493; sgat.
KW Transport; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 33 53 POTENTIAL.
FT TRANSMEM 57 77 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 282 302 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 357 377 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 420 440 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT CONFLICT 470 484 RAEEDAEKQLAQSOSA -> AQKKMKNNQNSLLNKEF
FT (IN REF. 1).
SQ SEQUENCE 484 AA; 52950 MW; 2CBBED0044BC6CAC CRC64;

Query Match 5.5%; Score 105.5; DB 1; Length 484;
Best Local Similarity 21.7%; Pred. No. 0.2;
Matches 72; Conservative 56; Mismatches 137; Indels 67; Gaps 14;

QY 10 NPPVLSQEFILQNHADIVSCVGMFEF---LGLVFEGTAEASIVFLTLQHSVAVPAAEQA 66
Db 37 NAPL-----LGIVTCLGYILLKRSVSVIIRKTIKIIFGMLLQAGSGILTSTFKP 87
QY 67 TGSKSLYYTGKVD-----LAT-----VFFYMLVAIIIIHATIQEYVLDKINK 107
Db 88 VYAKMSEVVGINGAISDTYASMMATIDRMGDAYSWSGYAVLLALNLN--ICVLLRRITG 145
QY 108 RMOFTKAKONKFNESQFSV--FYFSCIWGTFL-----ISENCSDPTLIWKA 155
Db 146 IRTIMLTGHIMFOOAGLIATVTLFYGYSMTTICTAILVSLYWGITSNNMKPT----- 200
```

QY 156 RHSMWTFQKFF--YISQAWWFAHFAPELXFOKTKKODIPQLVYIGLHLFH--ITGAY 211
Db 201 ---QEVTDGCGFSIGHQOQFASWI--AYKAPPLGKKEESVEDKLPGWLNIPHDNIVSTA 256
QY 212 LLYLNHGLGLLVLYFVELLSHMGCLFVFSDEKQKGLSWAIVFLLGRVLTIVSVLT 271
Db 257 IYMTIFFGAIL--LSFGIDTVQAMQVHMTVYILOTGFSFAVAIFITQGVMEVLAELS 314
QY 272 VGFHLAGSQRNPDALTGNNVLAAKIAVLSS 303
Db 315 EAFN-GISORLPGA-----VLAIDCAIYS 339
RESULT 12
ID S6A9 HUMAN STANDARD; PRT; 692 AA.
AC P48067;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium- and chloride-dependent glycine transporter 1 (GlyT1) (GlyT-1).
GN SLC6A9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=94239375; PubMed=8183239;
RA Kim K.-M., Kingsmore S.F., Han H., Yang-Feng T.L., Godinot N.,
RA Seldin M.F., Caron M.G., Giros B.;
RT "Cloning of the human glycine transporter type 1: molecular and
RT pharmacological characterization of novel isoform variants and
RT chromosomal localization of the gene in the human and mouse
RT genomes.";
RL Mol. Pharmacol. 45:608-617(1994).
CC !- FUNCTION: Terminates the action of glycine by its high affinity
CC sodium-dependent reuptake into presynaptic terminals. May play a
CC role in regulation of glycine levels in NMDA receptor-mediated
CC neurotransmission.
CC !- SUBCELLULAR LOCATION: Integral membrane protein.
CC !- ALTERNATIVE PRODUCTS: 3 ISOFORMS; GLYT-1A, GLYT-1B AND GLYT-
CC 1C (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC !- TISSUE SPECIFICITY: ALL GLYT-1 SUBTYPES CAN BE FOUND IN BRAIN,
CC KIDNEY, PANCREAS, LUNG, PLACENTA, LIVER. BUT GLYT-1C SUBTYPE IS
CC ONLY FOUND IN THE BRAIN.
CC !- SIMILARITY: BELONGS TO THE SODIUM:NEUROTRANSMITTER SYMPORTER
CC FAMILY (SNF).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S70609; AAB30784.1; -;
CC DR EMBL; S70612; AAB30785.1; -;
CC DR MIR; 601019; -;
CC DR InterPro; IPR000175; Na_neurotran_symport.
CC DR Pfam; PF00209; SNF; 1.
CC DR PRINTS; PR00176; NANEUSMPORT.
CC DR PROSITE; PS00610; NA-NEUROTRAN_SYMP_1; 1.
CC DR PROSITE; PS00754; NA-NEUROTRAN_SYMP_2; 1.
CC DR PROSITE; PS0267; NA-NEUROTRAN_SYMP_3; 1.
CC DR Neurotransmitter transport; Transport; Transmembrane; Glycoprotein;
CC Symport; Amino-acid transport; Transport; Alternative splicing.
FT DOMAIN 1 94 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 95 115 1 (POTENTIAL).
FT TRANSMEM 123 142 2 (POTENTIAL).

FT TRANSMEM 166 186 3 (POTENTIAL).
FT DOMAIN 187 273 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 274 292 4 (POTENTIAL).
FT TRANSMEM 301 318 5 (POTENTIAL).
FT TRANSMEM 354 371 6 (POTENTIAL).
FT TRANSMEM 383 404 7 (POTENTIAL).
FT TRANSMEM 437 456 8 (POTENTIAL).
FT TRANSMEM 485 503 9 (POTENTIAL).
FT TRANSMEM 519 539 10 (POTENTIAL).
FT TRANSMEM 560 579 11 (POTENTIAL).
FT TRANSMEM 598 616 12 (POTENTIAL).
FT DOMAIN 618 692 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 236 236 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 69 MAAAGPVAQSSPSQVTLTPVQRSEFLPFPFSGATPSTSLAE
FT SVLKVWHGAYNSGLLPQLMAOHSMAQ -> MYGKGAKGM
FT L (IN ISOFORM GLYT-1A).
FT VARSPLIC 16 69 MISSING (IN ISOFORM GLYT-1B).
SQ SEQUENCE 692 AA; 76823 MW; FABA3243A0D98073 CRC64;
Query Match 5.3%; Score 102; DB 1; Length 692;
Best Local Similarity 19.3%; Pred. No. 0.57;
Matches 64; Conservative 42; Mismatches 91; Indels 134; Gaps 14;
QY 26 IVSCVGMFFLLGLVFEQTAESIVFLTQHSVAVPAEAEQATGSKLSLYYGVKDLATVFF 85
Db 276 LLGCLGVSWL-----VVFCLIRGV-----KSSGVVVF-----TATPPY 310
QY 86 YMLVAIIHATIQEYVLDKIRM--QFTKAKQKNFESGQSFVFFSCIWGTFILIS- 142
Db 311 VVLTILFVRGVTLEGAFDGMVYLPQWDKILEAKVWGDAASQIFYSLACAMGLITMAS 370
QY 143 -----ENCLSDPTLWKARPHSMWTFQMKFFYISQALATWFAHFAPELXFOKTKKQDIPQL 197
Db 371 YNKFHNNCYRDSVII-----SITNCA-----TS 393
QY 198 VYIGLHLPHITGAYLLYLNHGLGLLVLYFVELLSHMGCLFVFSDEKYKGIS----- 251
Db 394 VVAGVVFISILG---FMANHLGVDV-----SRVADHGFGLAFVA---YPEALTLLPISP 441
QY 252 LMAIVF-----ILGRVLTLIV-----SVLTGFGHLAGSONRNP 285
Db 442 LWSLFFEFMLILGLGTQFCLELTIVTAIVDEGVNWILOKKTYYTLGVAVAG----- 494
QY 286 ALTGNNVLAAKIAVLSSCTIQAVVTNNLI 316
Db 495 -----FLLGIPLTQSAGIYWLLL 512
RESULT 13
YGN9_YEAST
ID YGN9_YEAST STANDARD; PRT; 802 AA.
AC P53121;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Hypothetical 90.8 kDa protein in HUL5-SEC27 intergenic region.
GN YGL139W OR G2812.
OC Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-749 FROM N.A.
RC STRAIN=S288C / FY1769;
RX MEDLINE=97197983; PubMed=9046099;
RA Voet M., Defoor E., Verhasselt P., Riles L., Robben J., Volckaert G.;
RT "The sequence of a nearly unclonable 22.8 kb segment on the left arm
RT chromosome VII from Saccharomyces cerevisiae reveals ARO2, RPL9A,
RT TIF1, MRF1 genes and six new open reading frames.";

```

RL Yeast 13:177-182(1997).
RN [2]
RP SEQUENCE OF 616-802 FROM N.A.
RC STRAIN=S288C / FY1679;
RA MEDLINE=96437978; PubMed=8840506;
RX Escribano V., Eraso P., Portillo F., Mazon M.J.;
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
RT cerevisiae chromosome VII reveals SPC27, SSMlb, a putative
RT S-adenosylmethionine-dependent enzyme and six new open reading
RT frames."
RL Yeast 12:887-892(1996).
RN
RP SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO YEAST YAL053W AND S.POMBE SPAC1F7.03.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X99960; CAA68223.1; -
CC EMBL; Z72661; CAA96851.1; -
CC EMBL; Z72660; CAA96850.1; -
CC EMBL; X92670; CAA63357.1; -
CC SCD; S0003107; XGL139W,
CC Hypothetical protein; Transmembrane.
KW TRANSMEM 3 23
FT TRANSMEM 170 190 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 406 426 POTENTIAL.
FT TRANSMEM 468 488 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 526 546 POTENTIAL.
FT TRANSMEM 558 578 POTENTIAL.
SQ SEQUENCE 802 AA; 90761 MW; 7BA13714AD912295 CRC64;

Query Match 5.28; Score 99; DB 1; Length 802;
Best Local Similarity 19.38; Pred. No. 1.2;
Matches 79; Conservative 53; Mismatches 134; Indels 144; Gaps 16;

Qy 19 ILQHADVSCVGMFFL-----LGLVPE-----GTEA----- 46
Db 146 ITNNVSEVLAQIQYFSGKVSQIGYKWTAVVAGIGLLLSAILSTFGNSTASHISAN 205

Qy 47 -----STVFTLQHSVAPV-----ABE-----QATGSKS 71
Db 206 TMSLFLYFQSVVYVAMQHVHPPIAAWAENLVMSGLIRISFMQIRFRWYVQSTGTP 265

Qy 72 LYYGVKDLATVF-----FYMVAIIHATIQEY-----VLDKINKRMQFTKAKQKENE 121
Db 266 SLYLTSTSMVLAORSQWYLMELPLIKRATNVLYGNANTLIFRGIKRYKMGIENTSIV 325

Qy 122 SQQSVFFVFCISGTFILISENCISDPTLI-----WKARPHSMWTFQ---MKFFYI 170
Db 326 CTGTFVFLCGVVLGFIIVFKCCVELATRLGWTQKARFWEFRQWRMLKGLLRIYI 385

Qy 171 SOLAYWFHAFPELYFQKTKKODIPQLVYIGLHLFHITGAYL-----LYL 215
Db 386 G-----FVQLTILSFWEFTERDSPAIVIACLFILLSCGLMLWAARVTFPARRSVALYN 440

Qy 216 NHLGLLLVLYFVLLSHMCGLFY--FSDEKYQKGLSLWAIIVFILRLVTLVSVLTVG 273
Db 441 NPAALL-----YGDVYVHLKYGFFTFNANHYWNWVILLSYIF-----VKSLVVG 486

Qy 274 FHLAGSNR-----NP---DALTGNNVLAAKIAVLSS 303
Db 487 FAQASGQTVLFMFILDLFYFVAIIYKPYLDPRNTNIMLIATVTVVNS 536

RESULT 14

```

```

CYB_PARTE STANDARD; PRT; 391 AA.
ID P15585;
AC DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cytochrome B.
GN COB OR CYTB.
OS Paramesium tetraurelia.
OG Mitochondrion.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramesium.
OX NCBI_TaxID=5888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=STOCK 51;
RX MEDLINE=90174913; PubMed=2308823;
RA Pritchard A.E., Sellhamer J.J., Mahalingam R., Sable C.L.,
RA Venuti S.E., Cummings D.J.;
RT "Nucleotide sequence of the mitochondrial genome of Paramesium."
RL Nucleic Acids Res. 18:173-180(1990).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15917; CAA34052.1; -
CC PIR; S07743; S07743.
CC InterPro; IPR000179; Cyt_b_b6.
CC Pfam; PF00033; cytochrome_b_n; 1.
CC PROSITE; PS00192; CYTOCHROME_B_HEME; FALSE NEG.
CC PROSITE; PS00193; CYTOCHROME_B_QO; FALSE NEG.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
Heme.
FT METAL 72 72 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 86 86 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 173 173 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 187 187 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 391 AA; 46044 MW; FCF34BAC4A543C25 CRC64;

Query Match 5.18; Score 98.5; DB 1; Length 391;
Best Local Similarity 21.58; Pred. No. 0.59;
Matches 79; Conservative 58; Mismatches 131; Indels 99; Gaps 20;

Qy 25 DIVSCVGMFFLLGLVFGSTAASIVFTLQHSVAVPAEQAQTSKSLYYGVKDLATVF 84
Db 17 EVFSLFGFTFTWITIVQLVSGTMLAFSSVPEPLIPTVRDEE-----DIEDLYTDD 67

Qy 85 FYMLVAIIHATIQEYVLDKINKRMQF---TKAKQNFN-----ESQFSVYFFFSKI 134
Db 68 FFWL-----HERGVDLIFISYFHLRLKLYLVNFDLETAESWKSQSVFLVQVWV 118

Qy 135 WGTFILISENCLSDPTLIWKARP-HSMMTFQMKFFIISQLAYWFHAFPELYFQKTKKODI 193
Db 119 FFGVLVLCCTH-LSEITLITIANIEHTFEMFKGK-----AYWF-----LFTDQLWTD 165

Qy 194 PRQLVYIGLHLFHITGAYLILYLNHLGLLLVLYFVLLSHMCGLFY-FSDEKYQKGIS- 251
Db 166 LIRLAYA-----HVSFAF--YLSFLGLL-----HGIDIHWDKNEPFYDGLSS 206

```

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 6, 2002, 17:20:26 ; Search time 80.19 Seconds
(without alignments)
796.048 Million cell updates/sec

Title: US-09-807-470-4
Perfect score: 1920
Sequence: 1 MGLRKKSTKNPPVLVSQEFIL.....VGVTNRVNDPPKREKSS 369

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhch:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1357.5	70.7	374	11	Q91V04	Q91V04 mus musculus
2	1251	65.2	373	13	Q902M0	Q902M0 xenopus lae
3	1129	58.8	369	13	Q902M1	Q902M1 brachydanio
4	820	42.7	371	13	Q902L9	Q902L9 xenopus lae
5	811	42.2	370	11	Q924Z5	Q924Z5 mus musculus
6	602.5	31.4	368	5	Q9W5C2	Q9W5C2 drosophila
7	602.5	31.4	1575	5	Q9W5C3	Q9W5C3 drosophila
8	597.5	31.1	368	5	Q9U1L3	Q9U1L3 drosophila
9	494	25.7	371	5	Q9U3P5	Q9U3P5 caenorhabdi
10	487	25.4	373	5	Q9XXK7	Q9XXK7 caenorhabdi
11	472.5	24.6	159	11	Q9CVJ6	Q9CVJ6 mus musculus
12	156.5	8.2	393	11	Q9D6J1	Q9D6J1 mus musculus
13	146.5	7.6	308	10	Q9M6A4	Q9M6A4 lycopersico
14	143	7.4	394	4	Q9HA82	Q9HA82 homo sapien
15	134.5	7.0	400	5	Q95RN6	Q95RN6 drosophila
16	127	6.6	310	10	Q9LDF2	Q9LDF2 arabidopsis

17	116.5	6.1	296	10	Q9LJK3	Q9LJK3 arabidopsis
18	116.5	6.1	297	10	Q9M6A2	Q9M6A2 arabidopsis
19	116	6.0	380	4	Q96G23	Q96G23 homo sapien
20	113	5.9	639	16	Q98KL8	Q98KL8 rhizobium l
21	112	5.8	380	11	Q924Z4	Q924Z4 mus musculu
22	111.5	5.8	360	5	Q9XWE9	Q9XWE9 caenorhabdi
23	111	5.8	3124	5	Q9CVB8	Q9CVB8 plasmodium
24	109	5.7	378	5	Q9Y1V7	Q9Y1V7 halocynthia
25	106.5	5.5	331	5	Q9XU34	Q9XU34 caenorhabdi
26	106	5.5	402	8	Q9T6M3	Q9T6M3 globodera p
27	104	5.4	600	17	Q9HKU1	Q9HKU1 thermoplasm
28	101.5	5.3	414	11	Q9D6K9	Q9D6K9 mus musculu
29	101.5	5.3	414	11	Q924Z3	Q924Z3 mus musculu
30	100.5	5.2	339	5	Q9W4Z3	Q9W4Z3 drosophila
31	100.5	5.2	614	16	Q31615	Q31615 bacillus su
32	100	5.2	474	8	Q9G8P5	Q9G8P5 naegleria g
33	100	5.2	642	4	Q9UN76	Q9UN76 homo sapien
34	99.5	5.2	507	16	Q92194	Q92194 rickettsia
35	98.5	5.1	308	10	Q9FZ69	Q9FZ69 arabidopsis
36	98	5.1	330	5	Q9GVA0	Q9GVA0 suberites d
37	98	5.1	716	2	Q46317	Q46317 carnobacter
38	97.5	5.1	210	8	Q9G8P4	Q9G8P4 naegleria g
39	97.5	5.1	262	5	Q9VH57	Q9VH57 drosophila
40	97	5.1	479	10	Q9FEL8	Q9FEL8 medicago tr
41	96.5	5.0	469	16	Q32204	Q32204 bacillus su
42	96.5	5.0	486	8	Q9G861	Q9G861 malawimonas
43	96	5.0	1123	10	Q9LYR7	Q9LYR7 arabidopsis
44	95.5	5.0	293	8	Q35716	Q35716 romanomeri
45	95.5	5.0	534	10	Q9SYA2	Q9SYA2 arabidopsis

ALIGNMENTS

RESULT 1
Q91V04
ID Q91V04 PRELIMINARY; PRT; 374 AA.
AC Q91V04;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TRAM1 (UNKNOWN) (PROTEIN FOR MGC:11724).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029764; AAK38167.1; -
DR EMBL; BC012401; AAH12401.1; -
SQ SEQUENCE 374 AA; 43039 MW; E6C65250F68E4393 CRC64;

Query Match 70.7%; Score 1357.5; DB 11; Length 374;
Best Local Similarity 69.5%; Pred. No. 2.4e-116;
Matches 260; Conservative 45; Mismatches 64; Indels 5; Gaps 2;

Qy 1 MGLRKKSTKNPPVLVSQEFILQNHADIVSCVGMFFLLGLVFEGTAEASIVFLQLQHSVAVP 60
Db 1 MAIRKSKNKPVLVSQEFILQNHADIVSCVGMFFLLGLVFEGTAEASIVFLQLQHSVAVP 60
Qy 61 AAEQATGSKSLYYGVKDLATVFFVLMVAIIHATIQEYVLDKINKRMQFTKAKONKEN 120
Db 61 ATEEQATGASLYHYGDKDLATVFLVLMVAIIHATIQEYVLDKINKRMHFFSKTHSKEN 120
Qy 121 ESGQSFVFFYFFSCINGTFTLISENCLSDPTLIWKARPMSMTFQMKFFYISOLAYWFHAF 180
Db 121 ESGQSFVFFYFFSCINGTFTLISENCLSDPTLIWKARPMSMTFQMKFFYISOLAYWFHAF 180

Db	121	ESQLSAFYLACVNGTFTLISENYISDPTILWRAPPHNLMTFQTKFFVISQIAYLWLHAF	180
QY	181	PELYFQRTKKQDIPROLQVYIGLHLFHITGAYLLYLNLHGLLLLVLHYFVELLSHMGCLFY	240
Db	181	PELYFQRTKKEDIPROLQVYIGLYFLFIAGIAYLLNLHGLLVLLVHYFVEFLFHSRLFY	240
QY	241	FSDEKYQKGISLWAIIVFILGRLVTLIVSVLTGFFHLAGSQNRNPDALTCNVNVLAAKIAV	300
Db	241	FSDEKYQKGSFSLWAVLFGVJGRLLTLLSVLTGFGIARAEKNOKLDFSTGNFNLAVRIAIV	300
QY	301	LSSSCTIQAVYVNLNLTLQIRWVEDSNITQASCMKKK----RSRSSKKRTENG-VGVTS	355
Db	301	LASICITQAFMMKKEINFQLRRWRHSAFAQPPVKKRPATVKGRRSSKRGKTENGVGNGTVTS	360
QY	356	NRVDCPPKREKKS	369
Db	361	NGADSPNRKREKKS	374
RESULT	2		
Q902M0			
ID	Q902M0	PRELIMINARY;	PRT; 373 AA.
AC	Q902M0;		
DT	01-DEC-2001	(TEMBLrel. 19, Created)	
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)	
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	
DE	TRAM1.		
OS	Xenopus laevis (African clawed frog).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;		
OC	Xenopodinae; Xenopus.		
OX	NCBI_TaxID=8355;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hartmann E.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY029534; AAK40302.1; .		
SQ	SEQUENCE 373 AA; 43121 MW; 62AF49E9859769DF CRC64;		

```

RESULT 3
Q90ZM1 ID Q90ZM1 PRELIMINARY; PRT; 369 AA.
AC Q90ZM1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRAM1.
OS Brachydanio rerio (Zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY029529; AAK40297.1; -
SQ SEQUENCE 369 AA; 41743 MW; A3EE6263E3165846 CRC64;

Query Match 58.8%; Score 1129; DB 13; Length 369;
Best Local Similarity 58.2%; Pred. No. 2.1e-95;
Matches 217; Conservative 64; Mismatches 84; Indels 8; Gaps

Qy 1 MGLRKKTKNPVLSQBFILQNHADIVSCVGMFFLLGLVFEGTAEASIVFLTLQHSVAVP 60
Db 1 MGRKKTKNPVLSHEFVIQNHADIVSCVAVFLLGLMFEITSKAVLFTVQYVNTIP 60

Qy 61 A--AEEQATGSKSYLYGVKDLATVFFYMLVAIIIIHATIQEYVLDKINKRMQFTKAKQNK 118
Db 61 ANGGEPEAT--VNYFVGLKDVATIFFYMLVAIIIMHAIQEYVLDKINKRMHFSKTKHSK 118

Qy 119 FNESSQSFVFFSCWGTITLISENCLSDPTLIWKRPHSMFTQMKFFYISQLAYWFH 178
Db 119 FNESSQLSAYFLSCLWGASILSENILSNPVSLWDGYPHTLMQFMQKFFYICQLGYWLH 178

Qy 179 APPELYFQRTKKDIPRQLYVYIGLHFLHTGAYLLLNHGLLLLLVLYFVELLSHMCGL 238
Db 179 AIPELYFQAKKEDIPRQLVVISLYLVHIAAGAYALNLRGLGLVLLVLYFVELIFHVSRL 238

Qy 239 FYFSDKYQKGISLWAIIVFILGRILVTILVSVLTGVGFHLAGSONRNPDLTGNVNLAAKI 298
Db 239 IYFSENERSGFTVMVYLVFLVGLRLTLLSVLTGVGFGLAGAEKQGLNALAEGSFNVLFVRV 298

Qy 299 AVLSSSTCIQAYVTNLTITLWLRVWSDSNQASCKMKKRSRS--SKKRTENGV--GVETSN 356
Db 299 TVLAACITQAFMMKKFNFQLRRWEQA--QQTLLKKSSSSKSKSKRANGVSGVSGAS 356

Qy 357 RVDCPPPKRKEKSS 369
Db 357 GADSPRARKKEKSS 369

RESULT 4
Q90ZL9 ID Q90ZL9 PRELIMINARY; PRT; 371 AA.
AC Q90ZL9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRAM2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Hartmann E.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY029535; AAK40303.1; -
SQ SEQUENCE 371 AA; 43548 MW; 582231A5286D130B CRC64;

```


DT	01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE	HYPOTHETICAL 46.4 KDA PROTEIN.
OS	Homo sapiens (human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
[1]	
RN	SEQUENCE FROM N.A.
RP	TISSUE=EMBRYO;
RC	Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA	Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA	Yamanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA	Nanamiya K., Iwayanagi T.;
RT	"NEDO human cDNA sequencing project.";
RL	Submitted (AUG-2000) to the EMBI/GenBank/DBJ databases.
[2]	
RN	SEQUENCE FROM N.A.
RC	TISSUE=LUNG CARCINOMA;
RA	Strausberg R.;
DR	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AK022151; BAB13972.1; -;
DR	EMBL; BC009828; AAH09828.1; -;
DR	InterPro; IPR001356; Homeobox.
DR	SMART; SM00389; HOX; 1.
KW	PROSITE; PS50071; HOMEBOX_2; 1.
KW	Hypothetical protein.
SQ	SEQUENCE 394 AA; 46398 MW; 67228829BDED2801 CRC64;

Query Match 7.4%; Score 143; DB 4; Length 394;
 Best Local Similarity 25.4%; Pred. No. 5.6e-05;
 Matches 46; Conservative 39; Mismatches 84; Indels 12; Gaps

Qy	108 RMQTTKAKQNKGESGFVFFSCITGWFILISENCLSDPLIKWARPHSMVTQMKF 167
	: : : : : : : : : : : : : : : : :
Db	126 RPQUTK-----KFCASRRFLFYLSSSFVGGLSVLYHESWLWAPYMCWDYRPNQTLKPSLYW 181
	: : : : : : : : : : : : : : : : : : : :
Qy	168 FYISQLAYWFHAPPELFQTKODIPRLQVLVIIGLHLFITGTGAULLNLHGLLLLVLYH 227
	: : : : : : : : : : : : : : : : : : : :
Db	182 WYLELGFGYLSLLIRLPFD-VKRDFKEQHIFHAVILMTFVSANLLRGSLVLLLHD 240
	: : : : : : : : : : : : : : : : : : : :
Qy	228 FVELLSHMCGIFYFSEDKYGKGIS----LMAIYFILGRVLTIVSVLTGVFHLAGSONRN 283
	: : : : : : : : : : : : : : : : : : : : : : :
Db	241 SSDVYLLEACKMVNYM--QYQVCDALPIFSFFYYTRVLVFPTQILYTYYESIS-NRG 297
	: : : : : : : : : : : : : : : : : : : : : : :
Qy	284 P 284
Db	298 P 298

RESULT 15

Q59RN6	PRELIMINARY;	PRT;	400 AA.
ID Q59RN6			
AC Q59RN6			
DT 01-DEC-2001 (TReMBLrel. 19, Created)			
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)			
DE D18904P.			
DN L18904P.			
GN CG15898.			
OS Drosophila melanogaster (Fruit fly).			
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC Ephydroidea; Drosophilidae; Drosophila.			
OX NCBI_TaxID=7227;			
[1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=Y, CN BW SP:			
RA Skapleton M., Broksstein P., Hong L., Agbayani A., Carlson J.,			
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,			
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,			

